

Schools at 10^{20} eV and beyond

A project that explores a research frontier, attracts high-school students of both sexes and diverse ethnicities and that can be scaled up to international level deserves not only celebration but also the €1-million support that it has just won.

Where do the highest-energy cosmic rays come from? Where did all the science students go? In the Netherlands, where physics undergraduate numbers are taking a dive, experimentally addressing the first question is hopefully countering the problem reflected in the second, thanks to an imaginative scheme in which universities and schools are collaborating in particle astrophysics.

The “high-school project on astrophysics research with cosmics” (HISPARC) seems to fire up the enthusiasm of school students in a way that few other schemes have done. And, encouragingly, this is just one entry among several in a competition, run by the Altran Foundation, that this year focused on innovative school projects. Announced this week as the winner, its technology and approach will within a year become accessible to schools in other countries.

With its first detectors set up in 2002, HISPARC focuses on the air showers produced when cosmic rays hit the upper atmosphere, each stimulating a cascade of secondary particles travelling in the same direction. The higher the energy, the larger the area at the ground over which secondaries arrive. At the highest energies seen so far — more than 10^{19} electronvolts — the secondaries can be detected by an array of detectors spread over 100 km² or so. Such showers caused consternation when detectors discovered them in the early 1990s; the numbers of ‘primaries’ hitting the atmosphere had been expected to tail off below such energies.

Initiated by physicists at the University of Nijmegen, a cluster of schools was chosen to host a network of detectors of these extraordinary events. The detectors — more sensitive than those that made the original observations — consist of a pair of plastic scintillators that detect light given off as secondary particles arrive. Events that trigger coincident bursts of light in both detectors, timed with the help of the

Global Positioning System, register within the network. Now tens of such detectors are being developed in schools in collaboration with several Dutch universities (see www.hisparc.nl/NL/english.shtml).

A key to the project’s motivational success, according to its coordinators, is that the schools become stakeholders by working with researchers in constructing the detectors. Importantly, those who are motivated include an even balance of both sexes and a healthy representation across ethnic groups. And that motivation could grow further as the networks develop into a system capable of having a scientific impact, pursuing rare events for which only a few detectors exist.

The project’s impact will be magnified by its success in winning this year’s Altran Foundation for Innovation Award. The prize money of €16,000 (US\$19,000) will no doubt be useful. But much more useful will be an unusual and commendable feature of the awards, whereby the foundation, based in a major consultancy organization (see www.fondation-altran.org), will also provide €1 million of support-in-kind over 12 months. This support is intended to ensure that the technology and software become cheaper to build and that the project spreads to schools in other countries.

The short-listed entries show that physics is not the only area in which imagination is being deployed to involve schoolchildren in the ideas of science. Measurements of climate change, the development of hands-on mathematical instruments modelled on historical originals, and the use of easy-to-make fuel cells are among the other projects dreamt up by universities and museums, and already in action. At a time when everyone is agonizing about the dearth of interest in science among schoolchildren, such projects should not only be noticed but also nurtured. Wherever science students have been disappearing to, here’s hoping that these schemes can encourage some of them back. ■

On with the show

Why scientists should support an artist in trouble.

For more than a decade, the Critical Art Ensemble, a US-based art cooperative, has used scientific tools to produce commentaries of the ways science and capitalism are shaping modern society. But has their latest act gone too far?

The ensemble, whose past theatrical roles have included a biotech firm and a cloning cult, recently assumed the trappings of a military germ-warfare research team. While teaching the history of germ research in the United States, they subjected their audience to a fake anthrax attack using real, but benign, microbes.

The show was meant to provoke public discussion, but it may have also provoked a federal anti-terrorism task force. As reported on page 690, a federal grand jury will convene on 15 June to decide whether Steven Kurtz, one of the group’s founders, has broken US bioweapons laws by possessing laboratory equipment and bio-warfare literature reportedly used in the ensemble’s performance.

On hearing this news, many researchers’ first inclination may be to shrug their shoulders. Since 11 September 2001, scientists have

worked overtime to ensure that their labs comply with federal bioterror laws. If an artist is running a mini-laboratory out of his home and reading up on bioweapons, it seems obvious that he would attract the authorities’ attention.

True enough, but the draconian way in which the prosecutors are acting is unwarranted. Supporters of Kurtz’s group say that it consulted regularly with scientists to make sure the shows were safe, and it is clear from the ensemble’s past performances that no real attack was planned. As with the prosecution of some scientists in recent years, it seems that government lawyers are singling Kurtz out as a warning to the broader artistic community.

Kurtz’s work is at times critical of science, but researchers should nevertheless be willing to support him (see www.caedefensefund.org/support.html to find out how). Art and science are forms of human enquiry that can be illuminating and controversial, and the freedoms of both must be preserved as part of a healthy democracy — as must a sense of proportion. ■





Akilia blow
Quartzite analysis
rocks claims for early
life on Earth
p688



Sporting chance
Experts take action to
quash contamination
of diet supplements
p689



Art critics
Bacteria leave
performance artist
facing grand jury
p690



Back in the dock
Russian scientist faces
retrial over espionage
allegations
p692

Scientists cry foul as Elsevier axes paper on cancer mortality

Meredith Wadman, Washington

Contributors to a medical journal are refusing to publish in an upcoming issue on health in the semiconductor industry unless a controversial paper, axed by the publisher, is included.

The paper in question discusses rates of cancer mortality among former employees of computer manufacturer IBM. But Elsevier, the publisher of *Clinics in Occupational and Environmental Medicine*, announced last week that it would not publish the paper on the grounds that it was not a review article.

"It was determined that it was original research and the format of it wasn't appropriate for *Clinics*, which is a review journal," says Eric Merkel-Sobotta, a spokesman for the Amsterdam-based publisher.

But the guest editor of the upcoming issue, and many of its contributing authors, say that this is not the real reason for the paper's removal. They allege, instead, that Elsevier is bowing to pressure from IBM not to publish it. Elsevier denies that pressure from IBM had any bearing on its decision.

The paper presents an analysis of mortality records for 32,000 IBM employees over 32 years. These records were originally submitted to lawyers by IBM last year during a court case in Santa Clara County, California, in which the company was sued by two former employees who alleged that their health had been affected by their jobs. Richard Clapp, an epidemiologist at Boston University, Massachusetts, and one of the paper's two authors, was an expert witness for the complainants and analysed the data supplied by IBM on their behalf.

Nature has obtained a copy of the study, which Clapp co-authored with Rebecca Johnson, an epidemiologist who runs Epicenter, a consulting firm in Circle Pines, Minnesota. It reports significantly more deaths from several kinds of cancer in IBM employees than would be expected from data for the general population, and higher-still cancer



A rash of lawsuits reflect concern over the safety of chip manufacture.

death-rates for workers who spent at least a month at IBM chip-manufacturing plants.

"If the material is worthy, and the editor is satisfied the piece meets the journal's scientific standards, the publisher's interference smacks of censorship," says *Clinics* contributor Daniel Teitelbaum, a medical toxicologist at the Colorado School of Mines in Golden. "This is precisely the problem presented by Elsevier's refusal to publish the Clapp-Johnson paper."

Standing firm

Joseph LaDou, the guest editor of the *Clinics* issue on the electronics and semiconductor industries, which is scheduled for publication in November, adds: "We are standing together with Clapp and Johnson. We will publish when they publish, wherever that may be." LaDou, who is director of the International Center for Occupational Medicine at the University of California, San Francisco, says that the paper had passed muster with four peer-reviewers drawn from among the issue's contributors.

He adds that on 6 June, the day that Elsevier informed him that the Clapp-Johnson paper would not be published, he called on the journal's contributors to consider withholding their pieces from publication. By 9 June, seven of the nine contributors had agreed to withhold their papers. As of 14

June, LaDou had not heard back from the other two. But Elsevier remains unmoved. "Boycotts are neither constructive nor effective," says Merkel-Sobotta.

Lawyers for IBM argue that the data used in the paper fall under a protective court order that prohibits their use for anything other than the Santa Clara lawsuit. "Dr Clapp cannot arrogate to himself judgements about what court orders have to be obeyed and what ones don't," says Robert Weber, a Cleveland-based attorney who represented IBM in the lawsuit.

LaDou had also agreed to be an expert witness in the lawsuit,

although the plaintiffs' lawyers did not call on him. The suit was decided in favour of IBM by a jury in February. The judge did not allow the data, and the conclusions drawn from them by Clapp and Johnson, to be presented to the jury, writing that they do "not and cannot demonstrate that there is any connection between the results of the study and exposure to chemicals at IBM". Some 200 similar lawsuits against the company are still in process.

But Clapp says that the broad pattern demonstrated by the huge database is important. He referred questions on the court order to his attorney, Indira Talwani. She says that IBM failed to mark as 'confidential' the deposition containing the study and Clapp's lengthy discussion of it with IBM lawyers. As a result, the plaintiffs' lawyers sent it to *The New York Times*, which published details of it last September. In addition, she says, any citizen may access the deposition containing the study at the Santa Clara County courthouse. "The stuff is in the public domain now, because IBM didn't take the right steps," she says.

Clapp initially withdrew the paper from *Clinics* in March, after being warned by an IBM lawyer not to publish it, Talwani says. But he resubmitted it after taking legal advice and concluding that IBM no longer had the legal right to keep the study private. ■

Funding review set to buck up basic research

Jim Giles, London

British scientists are looking forward to a spending announcement that could set the scene for a major revival in their fortunes over the next decade.

Britain's chancellor of the exchequer Gordon Brown is expected to announce plans to almost double spending on scientific grants and infrastructure over the next ten years, as part of a general spending review, which is provisionally scheduled to be released on 24 June.

The review is likely to favour established areas of priority, including clinical studies and spending on climate-change research. "There are bound to be winners and losers," notes Ian Gibson, Labour member of parliament for Norwich North and chair of the House of Commons science and technology select committee.

A report in *The Times Higher Education Supplement* on 11 June said that the plan would increase spending on the research councils, which fund most basic research grants in Britain, by 5.7% a year for ten years. A senior official close to the process says this is "not out of line" with what the final review will contain.

The official added that the review will emphasize basic scientific research, which the Labour government believes to be an important competitive advantage for Britain.

The plan will also make special provision for the support of interdisciplinary research, the official said, and could lay the ground-



Best case scenario: Gordon Brown's plan could double spending on science over the next ten years.

work for scientists and engineers at British universities to supplement their salaries through research income, as is possible in the United States. Inequalities in academic salaries are seen as a significant contributor to the brain drain of British scientists across the Atlantic.

Brown's plan was warmly welcomed by Peter Cotgreave, director of Save British Science, a lobby group formed at the height of that brain drain in the early 1980s. "It's

great that they are addressing the issues we have been highlighting," he says.

The Labour government has already increased the budget of the research councils by £1 billion (\$1.8 billion) to £2.2 billion since it came to power in 1997. But Cotgreave says that researchers continue to struggle to win grant funding, partly because so much of the investment has gone on infrastructure. "There is still no better chance of getting basic research funded," he says.

If it is implemented, next week's plan could change that. But the money is unlikely to be spread evenly between the seven research councils; medical and environmental research are likely to fare best.

A special fund for interdisciplinary projects is expected to be formed under the strategy and to be administered by Keith O'Nions, an earth scientist who became director-general of the research councils in January.

The question of academic salaries is also set to be addressed in next week's announcement. The government is thought to be considering a scheme that would allow university researchers to receive their current annual salaries from universities for nine months' work and to earn supplemental income from research grants during the rest of the year. But the success of such a scheme would depend on the willingness of independent grant-funding bodies, such as the Wellcome Trust, to support salaries as well as research. ■

Fresh study questions oldest traces of life in Akilia rock

Rex Dalton, San Diego

A famous study that found hints of the earliest life on Earth in Greenland rock has been thrown into doubt after researchers failed to replicate the work.

The quartzite rock, which comes from Akilia Island, was reported to contain organic carbon in crystals of a bone-like material called apatite — a promising sign of life (S. J. Mojzsis *et al.* *Nature* 384, 55–59; 1996). The rock was dated at about 3.85 billion years old, millions of years older than other signs of life found in Greenland and hundreds of millions of years older than evidence of life elsewhere.

The findings have been disputed before. Some researchers say the carbon cannot be proven to have a biological origin (C. M. Fedo and M. J. Whitehouse *Science* 296, 1448–1452; 2002). Others have suggested that the rock is not as old as originally thought (Y. Sano *et al.* *Nature* 400, 127; 1999).

There is now a fresh challenge. On



Promising the Earth: do these Greenland rocks, billions of years old, contain organic material?

11 June at a Goldschmidt Geochemistry conference in Copenhagen, Denmark, Aivo Lepland told scientists that he could not replicate the original study. Lepland, now with the Geological Survey of Norway, Trondheim, and his colleagues spent five years examining 15 different rock samples from the same outcrop in Akilia, but found no organic carbon in apatite crystals.

"I think there must have been a mix-up of samples," says Gustaf Arrhenius, a geochemist at the Scripps Institution of Oceanography in La Jolla, California, who is the senior author on both the original report and the recent presentation. Arrhenius says a sample from Greenland's nearby Isua formation, where there are organic carbon crystals in younger rock, may have inadvertently been attributed to Akilia. The original study may also have hit upon a genuine but rare find.

The original Akilia rock analysis was performed by Steven Mojzsis, then Arrhenius's doctoral student and now a geochemist at the University of Colorado in Boulder. Mojzsis, who attended the conference, told *Nature* he doubted there was a mix-up of samples, but he declined to discuss the specifics of Lepland's work. He and Lepland have agreed to divide the original samples for more analysis and have gone to examine Akilia and other sites in Greenland this week. ■

Dutch set the pace in bid to clean up diet supplements

Alison Abbott, Indianapolis

Nutritional supplements are riddled with contaminants such as steroids and other banned substances, scientists heard at the annual meeting of the American College of Sports Medicine (ACSM) in Indianapolis on 2–5 June. But a new initiative in the Netherlands could serve as a model for countries that want to give their athletes a safer choice, they said.

Supplements contain ingredients such as vitamins, minerals and amino acids. But for several years contaminants have been suspected of causing some athletes to test positive for banned substances, leading authorities to recommend that sportsmen and women stay off them altogether.

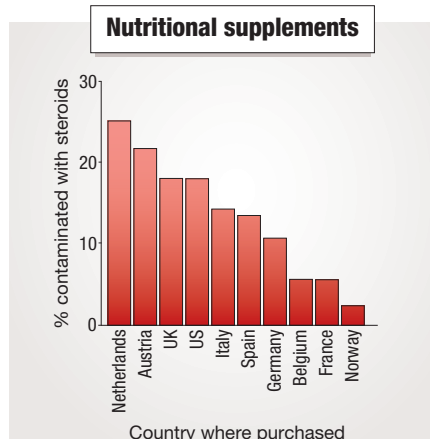
A recently published study, conducted at the German Sport University in Cologne, tested 634 supplements bought in 13 different countries between 2000 and 2001 (H. Geyer *et al. Int. J. Sports Med.* 25, 124–129; 2004). Although their labels claimed they were hormone-free, nearly 15% of the supplements tested positive for anabolic steroids (see Graph). The authors say that this could be the result of cross-contamination from improperly cleaned machinery, rather than intentional spiking.

In 2002, a study by the Netherlands Centre for Doping Affairs found that nearly a quarter of samples from supplements that Dutch athletes intended to use while training for the Utah Olympic Games were likewise contaminated.

Given that athletes continue to use supplements, several countries have taken steps to catalogue clean sources. Last year, for example, the German Sport University tested 200 supplements approved for medicinal use and found that none contained detectable levels of steroids. "This is clearly down to good manufacturing processes," says Wilhelm Schänzer, head of the Cologne Anti-Doping Institute, who is planning wider studies.

The biggest push to clean up supplements is going on in the Netherlands, the conference was told. The NZVT (Netherlands Security System for Nutritional Supplements in Elite Sports), launched last November, sets manufacturing standards for participating companies — of which there are 25 so far. It then tests each batch of supplement and posts the results online. Random tests are also carried out at later dates. More than 125 batches have been tested so far, all of which are listed on the website as clean.

Scientists at the meeting were impressed by such moves. But they remained concerned that athletes see supplements as performance boosters, despite the lack of evidence for this.



Tennis star Greg Rusedski was cleared of drug charges earlier this year after he said that he had taken contaminated supplements unwittingly.

It has never been proven that supplements do more good for an athlete than a balanced diet, said Ron Maughan, a nutrition expert from the University of Loughborough, UK.

Exercise psychologist James Whitehead from the University of North Dakota, Grand Forks, was one of many who complained that some exhibitors at the meeting were pushing supplements. "It has been a problem here for years," he said. Karl Fischer, who will soon become president of the ACSM, said that the claims of advertisers are checked by the college for sexism and slander but not for scientific content. "But this will change," he promised.

www.necedo.nl

Genomics institute rejects Venter's streamlining plan

Erika Check, Washington

Chiefs at The Institute for Genomic Research (TIGR) in Rockville, Maryland, and affiliated organizations led by bioentrepreneur Craig Venter have rejected a plan to consolidate their workforces — at least for now.

Venter co-founded TIGR with his wife, Claire Fraser, in 1992. In 2002, he started three other organizations using funding from the J. Craig Venter Science Foundation: the Institute for Biological Energy Alternatives, the Joint Technology Center and the Center for the Advancement of Genomics. TIGR is partly funded by the foundation, but it also receives money from outside sources such as the National Institutes of Health.

Earlier this year Venter proposed combining the workforces of all of these organizations. "We have five not-for-profit organizations and five boards and we have duplicate human resources, finance and other departments, and it's been draining the money that comes out of the foundation," Venter says.

Venter is chairman of TIGR's board, whose membership overlaps with those of the foundation and the other institutes. Fraser has directed TIGR for the past six years. But there are concerns that the present set up might lead to tension because Fraser and Venter, who have been married since 1981, are now separated.

Venter says that this did not influence his decision to seek consolidation. "Both myself and the board have been very supportive of her job," he says of Fraser.

Late last month, TIGR's board decided to keep the institute separate from the other three organizations, which will merge into a single entity. "TIGR really has a different style and mission from the other organizations," says Gerald Rubin a TIGR board member and vice-president of the Howard Hughes Medical Institute in Chevy Chase, Maryland.

But Venter did not rule out future reorganizations. "Nothing is ever final," he says.



Apart ... but still working together: Craig Venter (left) and Claire Fraser.

M. THEILER-FILES/REUTERS (LEFT); TIGR

Future brightens for Eddington project to hunt distant planets

Laura Nelson, London

On the same day that Venus was observed as a black disk traversing the Sun, the European Space Agency (ESA) was advised to resurrect a mission based around such mini-eclipses. If built, the Eddington space-based observatory will use similar events to search for planets in other solar systems.

The €210-million (US\$250-million) project would consist of three telescopes arranged to track the light intensities of many stars over several years. The plan is to place the structure beyond the Moon to allow uninterrupted observation of thousands of stars.

When, as viewed from Earth, Venus crossed the path of the Sun on 8 June, the silhouette of the planet reduced the star's brightness by a tenth of a per cent. Eddington is designed to detect changes a thousand times smaller than that, and will use dips in star brightness to deduce the existence of distant planets.

Eddington will also investigate what the insides of stars are made of, by measuring the frequency of emitted light and deducing information about their chemical and physical properties. This may illuminate star evolution: "Studying younger and older stars will help us understand the Sun," says Ian Roxburgh, an astronomer at Queen Mary, University of London, and a coordinator in the mission's design team.

The project was approved in 2002 for launch in 2008, but was shelved last November after its design phase, because the ESA budget would not stretch to cover its cost.

On 8 June, ESA's Space Science Advisory Committee (SSAC) recommended that the project be given precedence after receiving advice from several other committees. "The SSAC strongly endorsed Eddington," says Giovanni Bignami, who is chairman of the committee and director of the CESR, the laboratory of space astrophysics in Toulouse, France. "It is our first priority."

Eddington now tops the space agency's list of unfunded science missions. "If ESA receives more money, Eddington would be the priority," says Roxburgh.

Jean-Jacques Dordain, director-general of ESA, is currently seeking an extra €700 million in funding for the agency from its member states. Eddington should be built if a fraction of that comes through. ■



Taking the strain: agents took lab equipment from artist Steven Kurtz's home following his wife's death.

Bacteria raid may lead to trial for artist tackling biodefence

Geoff Brumfiel, Washington

An artist who uses live organisms and laboratory equipment in his performances is being scrutinized by federal officials who suspect he has broken bioweapons laws.

Federal prosecutors are deciding whether to file charges against Steven Kurtz, an art professor at the State University of New York at Buffalo, for possessing strains of bacteria, laboratory equipment, and several books on biowarfare. But Kurtz's friends and supporters say that the materials are part of his group's performance pieces.

Since 1987, 46-year-old Kurtz has been part of a Buffalo-based performance group, the Critical Art Ensemble, whose work offers political commentary on scientific topics. They often incorporate laboratory techniques into their shows, according to Nato Thompson, a curator at the Massachusetts Museum of Contemporary Art in Boston who has seen the group's work. "The lab experiment is the performance," he says.

The current investigation began with the death of Kurtz's wife, Hope, on 11 May. Medical staff responding to an emergency call from Kurtz noticed some "materials that made them uneasy" in his home, according to Paul Moskal, a spokesman for the Federal Bureau of Investigation. Officers from the bureau searched the house and questioned Kurtz about materials found there.

Moskal declined to comment on what the search uncovered, but Claire Pentecost, a photographer at the Art Institute of Chicago and long-time friend of Kurtz, says that investigators carried off samples of *Bacillus globigii*, *Serratia marcescens* and a benign

form of *Escherichia coli*. They also seized laboratory equipment, including a polymerase chain reaction (PCR) machine and books on biowarfare, she says. The books were part of the group's latest project, *The Marching Plague*, which simulates an anthrax attack as part of its critique of government germ-warfare research.

Analysis showed that the biological materials removed from the Kurtz home posed no threat, and an autopsy of Hope Kurtz revealed she had died of a heart attack. Authorities have nonetheless decided to bring a case against him before a grand jury; this will determine whether he can be prosecuted in court.

The US attorney's office in Buffalo declined to comment on the case, but a website that supports Kurtz's group says the case is being pursued under the Biological Weapons Anti-Terrorism Act of 1989. The act was amended under the USA Patriot Act of 2001 to allow the prosecution of "whoever knowingly possesses any biological agent, toxin, or delivery system".

Jonathan Turley, a lawyer at George Washington University in Washington DC, says that he thinks prosecutors are unjustified in pursuing criminal charges. Turley defended Texas Tech University plague researcher Thomas Butler in his recent federal trial (see *Nature* 426, 593; 2003). "The clear expectation of Congress was that the justice department would exercise a modicum of judgment," he says. "This law is designed to deal with the likes of al-Qaeda, not Andy Warhol."

Kurtz was due to appear before a grand jury in Buffalo on 15 June, when his supporters planned to rally in his defence. ■

Stem-cell library boosts the case for change

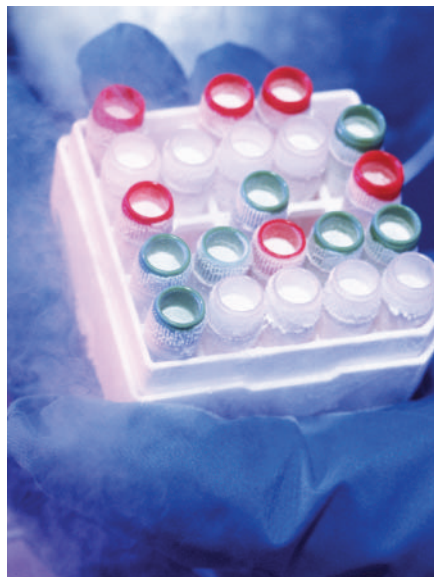
Erika Check, Washington

A fertility clinic in Chicago, Illinois, has announced that it has created a large library of embryonic stem-cell lines — many of them carrying disease genes.

The announcement is significant, stem-cell specialists say, because cell lines carrying disease genes are in short supply and could be useful for studying illnesses with a genetic component. The existence of a large number of such cell lines could increase pressure on President George Bush to change his stem-cell policy, which prohibits federally funded scientists from working on them.

Yury Verlinsky, director of the Reproductive Genetics Institute in Chicago, revealed the existence of the lines at a meeting of the International Society for Stem Cell Research last week. His colleague, Anver Kuliev, says that the group, which has not yet published details, has derived “a few dozen” human embryonic stem-cell lines from normal patients. They have also derived about a dozen lines that carry mutated disease genes including those for β -thalassaemia, a condition characterized by defective blood cells, and various chromosomal abnormalities.

Kuliev says his group has become skilled at producing embryonic stem-cell lines, and



Bright idea: derivation of extra stem-cell lines builds pressure on President Bush's restrictions.

has access to a unique patient population, because the institute runs a large pre-implantation diagnosis programme. This means that couples attending the clinic for assisted-reproduction services can have their

embryos screened for genetic diseases before they are implanted in the mother's womb. If the screening turns up disease genes, says Kuliev, stem cells are collected from the embryos, with the couple's consent, and propagated into stable lines.

“Our emphasis is on establishing embryonic stem cells with different genetic abnormalities,” Kuliev adds.

“I think this is highly significant,” says George Daley, a cell biologist at the Children's Hospital in Boston. “This is a large number of new lines, many representing models of human disease.”

Kuliev says that his group will make the newly created lines available to other researchers, although the terms of these collaborations are still to be worked out.

No scientist using federal funding will be able to use the cells, because President Bush's ruling of 9 August 2001 bars them from using lines created after that date.

This policy has been coming under pressure lately. On 4 June, 58 senators sent a letter to President Bush asking him to revisit it. And on 12 June, Senator John Kerry (Democrat, Massachusetts), who is his opponent in this November's election, took up the issue as part of his campaign. ■

K. GULDBRANDSEN/SPL

Alcohol report directs drinkers to their doctors

Helen Pearson, New York

A drink a day keeps the doctor away from some of us, concludes a review by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) in the United States. But you need to ask your doctor if your level of drinking is healthy, it adds.

Many people assume that a tipples or two a day is a tonic, but researchers believe this assumption is based chiefly on misleading media reports and wishful thinking. The institute reviewed scientific literature showing the relative risks and benefits of drinking in moderation (L. Gunzerath, V. Faden, S. Zakhari & K. Warren *Alcohol. Clin. Exp. Res.* 28, 829–847; 2004).

Rather than producing an easy rule of thumb, the review concludes that the healthiest level of consumption depends on an individual's age, sex, overall health and lifestyle — a calculation best totted up in the doctor's office. “We cannot give a blanket statement that alcohol is good for you so go out and enjoy,” says the institute's Samir Zakhari, one of the authors of the report.

The case for alcohol is based on findings that moderate drinkers are at lower risk of heart disease than either abstainers or heavy imbibers. The review weighs these benefits



Your very good health — as long as you are a moderate drinker over the age of 40.

against alcohol's myriad detrimental effects, such as the increased risk of injury, breast cancer, liver disease and other conditions.

When both benefits and risks are factored into the equation, the reviewers conclude, people still seem to reduce their overall health risks by having one or two alcoholic drinks a day. But the degree of benefit varies from person to person, says Zakhari. People aged 40 or more, who are already under some threat from heart

disease, seem to benefit the most, whereas younger drinkers may gain very little.

The NIAAA review adds that there is little consensus among experts about what constitutes ‘moderate’ drinking. Some studies consider this to be one drink a week, others as many as four a day.

The review's contents are likely to provide the basis for the alcohol section of new dietary advice due to be released by the US government next year.

Despite its complicated conclusions, experts welcomed the attempt to weigh up the pros and cons of moderate drinking. “It's as good as it could have been,” says Kenneth Mukamal, who studies alcohol and health at Harvard Medical School, Boston.

But some researchers doubt that the report's contents, and other reliable advice on healthy drinking, are reaching those already at the bar (see *Nature* 428, 598–600; 2004). Rather than disseminating a vague notion of moderation, drinks labels and trained bar staff should advise people not to exceed one or two drinks a day, suggests alcohol-policy expert Thomas Babor at the University of Connecticut, Farmington. “It's now a question of getting the message out there,” he says. ■

IMAGESTATE/ALAMY

Live anthrax bacteria inadvertently sent to vaccine researchers

San Francisco A batch of live anthrax seems to have slipped through several tests that should have ensured the bacteria were dead. The tests were done both by the company that supplied the anthrax and by the lab that used it to make a trial vaccine.

The bacteria (*Bacillus anthracis*) were thought to be dead following supposed heat inactivation and tests showing that they did not grow in culture. Researchers at the Children's Hospital and Research Center at Oakland in California then used the bacteria to create an anthrax vaccine. The bacteria were discovered to be alive when 49 of 50 mice inoculated with them quickly died. Seven people at the lab have been put on a precautionary course of antibiotics.

The Centers for Disease Control and Prevention in Atlanta, Georgia, and the supplier — the Southern Research Institute in Birmingham, Alabama — are investigating. The California Department of Health Services says that there is no risk to anyone else in the research building or in the surrounding community.



Under fire: Valentin Danilov faces a retrial over accusations that he passed secrets to China.

Retrial ordered for Russian physicist accused of spying

Moscow The Russian Supreme Court has overturned the acquittal of a Russian physicist accused of spying.

Valentin Danilov, a physicist formerly at Krasnoyarsk State Technical University in Siberia, was arrested in 2001 for allegedly supplying a Chinese company with classified satellite-technology information (see *Nature* 424, 477; 2003). He was found not guilty in December 2003. But on 9 June the high court ruled that Danilov's attorneys had improperly influenced the jury, and overturned the decision. His retrial has yet to be scheduled.

Kangaroo genome gears up for giant leap

Sydney Australia's wildlife is to get its day in the sun, after the kangaroo lost out to the opossum last year as the first marsupial to have its genome sequenced (see *Nature* 425, 753; 2003).

The US National Human Genome Research Institute last week announced a partnership with the Australian Genome Research Facility to sequence the tammar wallaby (*Macropus eugenii*, right), a small kangaroo whose genome is roughly the same size as that of a human. The project, beginning this year, should take about two years to complete.

"The evolutionary distance between opossum and wallaby is as great as that between human and mouse," says Jenny Graves, a marsupial geneticist at the Australian National University in Canberra, who led the push to have the kangaroo genome sequenced. Comparing the two should provide a wealth of information, she says.



Eminent scientists in Russia support Danilov's claim that the information he allegedly passed on was not secret, and say the case is part of a witch-hunt against scientists by the Federal Security Service, the main successor to the Soviet KGB. "They are fabricating an abomination against this person," says Vitaly Ginzburg, winner of the 2003 Nobel Prize in Physics.

Physicist Schön stripped of doctorate

Munich The University of Constance in Germany has withdrawn the PhD of Jan Hendrik Schön, the German physicist who fabricated data in 16 high-profile papers produced during his stay at Bell Laboratories in Murray Hill, New Jersey.

The university awarded Schön his PhD in 1997 for his work on semiconductor systems for use as solar cells. After he was found guilty of scientific fraud in 2002 (see *Nature* 419, 419–421; 2002), the university carried out an investigation of his earlier research in Germany, which was found to be free of misconduct.

But the university has now used a law that allows a PhD to be withdrawn if its holder behaves in an "undignified" manner. "Schön has violated the dignity of the doctorate, and thus damaged the credibility of science in public," says Wolfgang Dieterich, a physicist in charge of PhD awards at the university. Schön has one month to appeal the decision.

Radiation monitor gives early warning of pest attack

Tokyo Japan is turning its system for monitoring nuclear radiation on insects.

In the wake of the Chernobyl disaster, the Japan Atomic Energy Research Institute developed computer programs to model the

diffusion of radioactive substances. But researchers soon realized that the simulations could be used to model any number of things blowing in the wind, including volcanic ash — and the brown planthopper (*Nilaparvata lugens*). This tiny enemy of the rice farmer blows over from southern China during June and July. The 4-millimetre-long insect sucks the sap out of rice stems and can cause complete crop loss.

Starting this year, with help from researchers at the National Agriculture and Bio-oriented Research Organization in Tsukuba, the program will use wind and temperature data from global weather forecasts to predict up to two days in advance where and when the insects will land in Japan. Farmers can then use these predictions to limit their use of pesticides.

Hospital names head of breast cancer institute

Washington Cancer biologist Tak Mak has been made head of a new breast-cancer centre at the Princess Margaret Hospital in Toronto, Canada.

The Institute for Breast Cancer Research, which will initially house 12 faculty members, will work on all stages of cancer drug discovery, from target identification to clinical trials. It has a start-up fund of Can\$60 million (US\$44 million) in private and government money, and administrators hope to raise another Can\$65 million over the next five years, Mak says.

Mak was previously director of the Amgen Institute — a mouse genetics screening facility at the University of Toronto. Amgen withdrew funding from this venture two years ago (see *Nature* 417, 4; 2002), at which point Mak became director of the Advanced Medical Discovery Institute at the university. He will retain this position.

Every last drop

The price of petrol is going up and new oil discoveries are declining. Can underground fires and hydrocarbon-hungry bacteria keep the oil flowing? Jim Giles finds out.

Cheap oil is on the way out. This has been made abundantly clear in recent months thanks to security problems in Iraq and terrorist attacks in Saudi Arabia — both countries with massive crude-oil reserves. Markets are unsettled and prices at the pump are soaring. Last month the price of crude reached more than US\$42 a barrel, its highest for 20 years.

But geologists have known for years that the end of cheap oil is in sight. Since the 1960s, the rate at which new wells have been discovered has been on the wane¹. Forty years ago, more than 50 billion barrels of oil could be discovered in a single year. Today, finds of about 10 billion barrels per year are far more common (see graph), many of them from smaller fields. In short, many experts think that all the big gushers — the wells that spew out masses of oil cheaply and easily — have probably been found.

The decline has hit at least one oil company hard. Earlier this year, the Anglo-Dutch oil giant Shell was caught in a media storm after it admitted that its proven reserves — the oil the firm knows it can definitely extract — are 20% less than its previous estimates. Reserve estimates are often optimistic, analysts say, and many wells don't quite produce as much oil as expected. But new wells are usually found, making up the numbers. Analysts suggest that in this case Shell didn't find enough new wells to make up their accounts, drawing attention to the embarrassing shortfall.

Shell's misfortunes and the situation in Iraq may change. But these effects are overlying a trend that cannot be ignored. Oil companies now realize they cannot rely on finding new fields in the long term. Instead, they will need to extract every last drop they can from existing sources.

A slew of speculative techniques, from lighting fires in oil fields to using microbes to help wash out oil reservoirs, are already available to do this — and more are on the way. Some methods have been in development for decades. But only now is the price of oil becoming high enough to make them commercially viable.

"In the United States alone there is 355 billion barrels of oil that is not recoverable using existing techniques," says Betty Felber, a senior petroleum scientist at the National Energy Technology Laboratory in Pittsburgh, Pennsylvania. This is more than 15

times as much as the proven reserves that we can get at using conventional methods. "More and more people are applying these technologies," she says.

"We say oil is running out, but 60% of it is left in reservoirs," adds Malcolm Greaves, a geophysicist at the University of Bath, UK. "What are we going to do? Walk away?"

Greaves is pioneering one radical solution: setting fire to the reservoir. The technique was first trialled, albeit accidentally, in Russian oil fields about 50 years ago. Engineers pumped air into reservoirs in a bid to raise the pressure and force oil out of existing wells, and discovered that the air reacted with and ignited the oil. The combustion, which was limited to a small area of the reservoir, heated the oil, reducing its viscosity and allowing more of it to flow smoothly out of nearby wells.

The method, known as *in situ* combustion, has since been tested more rigorously, with mixed results. Greaves estimates that about 140 pilot projects were run during the 1980s, when oil prices were also high. Some suffered from 'blow-backs' — explosions that travelled back up the well through which the air was injected into the reservoir. In around a third of the experiments, oil flow was not increased as much as anticipated, in part because the injection well was often too far away — typically hundreds of metres — from the producing well.

Fire down below

The technique could be rehabilitated if a large-scale trial set to start this December is successful. Under scrutiny is a variant of *in situ* combustion known as toe-to-heel air injection (THAI), developed by Greaves and his colleagues at the University of Bath². As the wells are being drilled especially for the trial, they are designed to be just a few metres apart — avoiding some of the problems of previous trials. Another new feature of THAI is that the oil is drawn off through a pipe that runs horizontally, rather than vertically, through the reservoir. This means that the fire can move along the pipe, says Greaves, pushing oil out in front of it.

Greaves reckons that the US\$30-million trial at the Christina Lake oil field in British Columbia, Canada, could recover around 80% of the billion or so barrels of oil in the field — an ambitious estimate given that no more than 60% of a reservoir's field can normally be



Sky high: pump prices soar in the United States.

extracted. The THAI tests, run by the Canadian company Petrobank, of Calgary, should also prove useful in extracting Christina Lake's more viscous oil, made up of heavier, larger molecules. The combustion will split the oil, allowing the lighter and more valuable components to flow out of the reservoir.

For fields that already contain lighter oil, Egil Sunde has an alternative³. A marine biologist by training, Sunde has worked for Norwegian producer Statoil for around 20 years. "My idea was to use Nature's ways," he says. Some microbes feed naturally on hydrocarbons. These bacteria are already used to help strip oil from polluted beaches, he points

J. GRESS/REUTERS/CORBIS



Field of dreams: new technologies could pull hidden reserves from old wells.

out, so perhaps they could also help pull it out from underground reservoirs.

Sunde is using microbes that both feed on oil and make it less sticky. Oil is difficult to extract from half-empty fields, in part because it clings to the pores in the reservoir rocks. Water can be pumped in to help push out the oil, but eventually it will simply flow over the top of this residue.

Sunde's microbes grow at the interface between oil and water in the rock pores. This helps water molecules grab onto the oil and detach it. "This makes the oil move more easily through the pores," says Sunde, who for commercial reasons won't reveal the species of bacteria that do the trick best.

Statoil has been testing this idea since 1991 and was sufficiently encouraged to start a commercial-scale trial in 2001. Bacteria grown in Sunde's lab are now being pumped, together with nutrients and oxygen, into reservoirs in the Norne field off the coast of Norway. Sunde says it is too early to judge the results of this larger trial, but hopes that it will eventually increase the amount of recoverable oil — estimated to be around 530 million barrels — by about 5%.

The technique will not work in every oil field — the pores in the chalk reservoirs of many Middle East fields are too small for bacteria to pass through easily, for example. But industry observers are cautiously optimistic about the technique's potential in sandstone reservoirs like those in the North Sea. "This has a big future," insists Sunde. "It could revitalize thousands of fields."

So far there only a few techniques such as Sunde's and Greaves' are ready for large-scale trials, but other fledgling techniques could mature soon. "They will become appropriate now that the oil price is high," says David Hughes, principal reservoir engineer at Reservoir Management Limited in Aberdeen, UK.

High pressure

The UK-based energy company BP, for example, is investigating the chemistry that governs the reaction between water, oil and reservoir rock. Researchers there think that decreasing the salinity of the water could prevent oil drops from becoming trapped in the pores⁴.

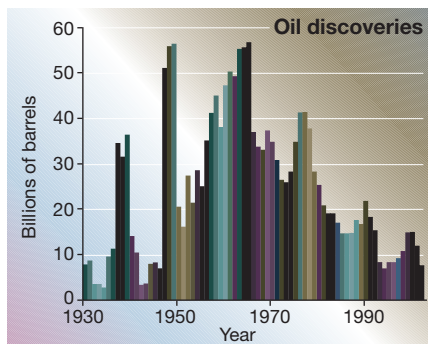
Other researchers think that plugging holes in the rock with plastic will make it easier to build up water pressure with injected fluid. Another way to increase the pressure and force out the oil could be to use gas-producing microbes. Even more speculative ideas are

being pursued, such as using microwaves to heat up the oil and lower its viscosity⁵.

But the enthusiasm needs to be taken with a pinch of salt. Oil companies tend to hush up unsuccessful projects, points out Nigel Brealey, also at Reservoir Management Limited. In the early days of *in situ* combustion projects, he says, the fire sometimes broke through into other wells and caused explosions. "But it wasn't well reported," adds Brealey, "People don't talk about failures." Techniques that look good on paper can also fail to make economic sense in the field. Despite years of work, unconventional methods such as microbial extraction produced some 2.8 million barrels of oil per day in 2003 — just 3.5% of the global total.

That percentage will, however, almost certainly rise as the price of oil increases. The techniques will not make oil cheaper, but they will help keep it flowing as new discoveries dry up. The days of easy oil extraction may be over. But while there is money in it, techniques to pull the last drop of the black stuff from the ground are going to be in demand. ■

Jim Giles is a reporter for Nature, based in London.



Asia's tigers get the blues

Mental illness, in particular depression, seems to be on the rise throughout East Asia. Why is this so? And can psychiatrists and the pharmaceutical industry together turn back the tide? Carina Dennis investigates.

In China, old people are killing themselves at an unprecedented rate. Once venerable and respected, the elderly have been left feeling uncertain of their status as rapid economic development has transformed family and community structures. Many give up hope, and end their lives with a dose of agricultural pesticide. "The suicide rate among the elderly in parts of Asia is shocking," says Helen Chiu, a psychiatrist at the Chinese University of Hong Kong.

In Japan, meanwhile, roughly a million youths have turned their backs on society. Some lock themselves in their rooms to sleep all day and play computer games all night. This social phobia — known as *hikikomori* — is seen as a symptom of a young generation unable to cope with socio-economic uncertainties. The decades-long Japanese economic boom has collapsed, with the emergence of an unemployed underclass.

These distressing phenomena are just part of what seems to be a general upward swing in the rates of mental illness being recorded across East Asia. In part, the explanation lies with rapid societal changes that have swept through the region as many of its nations have flexed their economic muscle. But some experts argue that recent studies may also have unmasked long-standing issues with mental health that went unrecognized in societies that have traditionally tended to view psychiatric illness as a sign of moral weakness.

Whatever the explanation, many Asian nations are currently ill-equipped to provide the specialist care needed to rehabilitate those who are suffering from mental illness. East Asian people also tend to possess genetic profiles that make them respond very differently to the antidepressant and antipsychotic



drugs developed to help bring relief to Westerners. If the region's problems with mental health are to be tackled, psychiatrists, geneticists and the pharmaceutical industry all have a lot of work to do.

Severe mental illnesses, such as schizophrenia and other psychotic disorders, are relatively easily recognized and are thought to occur at similar rates in societies throughout the world. But when psychiatrists first cast their gaze East in the 1950s, they concluded that Asia was almost entirely free of the subtler mood disorders that are prevalent in most Western societies^{1,2}.

Even the most recent data gathered for the World Health Organization and published this month³ suggest that people in China and Japan suffer less from depression and related disorders than their Western counterparts. But some experts argue that these figures fail to reflect the true magnitude of the problem — particularly given that Asia has some of the highest suicide rates in the world (see Figure, opposite). "If we consider suicide as an extreme endpoint of depression and other mood disorders, it suggests that we cannot have less of a problem with mental illness than the West," says Sing Lee, a psychiatrist at the Chinese University of Hong Kong.

Nearly 300,000 Chinese kill themselves each year, mostly in rural areas⁴. It is one of

the few countries in which more women than men commit suicide — accounting for more than half of all female suicides worldwide⁵. Michael Phillips, a psychiatrist at the Beijing Suicide Research and Prevention Center at Hui Long Guan Hospital, suspects that the high suicide rate in China is in part due to attempted suicides being more successful than in the West. "I don't believe all those people wanted to kill themselves," he says. Rather, he argues that the highly toxic pesticides widely used in rural China for suicide attempts and scant resuscitation facilities lead to high fatalities from what might elsewhere be considered a cry for help.

Broken homes

Chiu links the apparent rise in suicide with the changing socio-economic climate across East Asia. This has brought diminished financial and status security, rising unemployment and the breakdown of traditional family and community networks. Large, tightly knit family groups have been replaced by nuclear families that move from their communities in search of work, often great distances away. And in China, the breakdown of family-support networks may be set to escalate, as the one-child policy, introduced in 1979 to control population growth, makes its effects felt.



Behind closed doors: in Japan, about a million young people have turned their backs on society.

At the same time, substance abuse, which often goes hand-in-hand with mental illness, is on the rise. And eating disorders, once thought to be exclusive to Western societies, are now emerging in cities such as Hong Kong, Beijing and Shanghai. "Until about 15 years ago, we didn't see anorexia patients. But now we see hundreds of young Chinese girls each year," says Lee, who runs an eating disorder clinic in Hong Kong.

Depression is still widely under-reported, however. "The vast majority of patients with mental illness in China go undiagnosed and untreated," says Arthur Kleinman, a psychiatrist at Harvard University Medical School, who has been investigating mental illness in China for more than two decades.

One reason is the dearth of doctors trained in mental health — there are thought to be only around 15,000 with proper qualifications for China's billion-plus people. And for many Chinese, standard techniques used to diagnose depression are of little use. "If you ask Western questions about depression

in an urban population, they will basically understand," says Phillips. "But if you go to remote rural areas and ask the same questions, they will have no idea what you are talking about."

Weak nerves

More generally, East Asian patients tend to describe their psychiatric problems in terms of physical symptoms. "A patient in the United States might say they feel depressed or have suicidal thoughts, whereas a patient in Asia is more likely to talk about chest pain, headaches and fatigue," says Chiu.

This is exemplified by the concept of 'neurasthenia', a condition widely diagnosed in China that literally means 'weak nerves'. In the early 1980s, Kleinman revealed that the majority of neurasthenic patients had symptoms that fitted the criteria for depression, and that their condition improved if they were treated with antidepressant drugs⁶.

Diagnosing patients as neurasthenic may be popular because it avoids the stigma of

labelling them as mentally ill. "Mental illness is typically not thought of as a treatable illness in Asia, but rather as a moral weakness and the result of a faulty upbringing," says Chiu. "It is associated with a lot of shame and guilt for both the individual and the family."

While China's problems with mental health illustrate the rising trend being seen across the region, experts caution against making generalizations for all of East Asia. Even within countries, clinical symptoms of mental illness can differ from city to city⁷. And cultural differences between Asian nations can lead to mental illnesses being expressed in markedly different ways. Severe anxiety is a case in point. *Shenkui* is a Chinese syndrome in which sufferers display panic symptoms arising from a fear of losing *yang* or positive energy, due to the loss of semen. *Koro*, exclusive to Malaysia, is a sudden and intense anxiety that the penis will disappear into the body and possibly cause death.

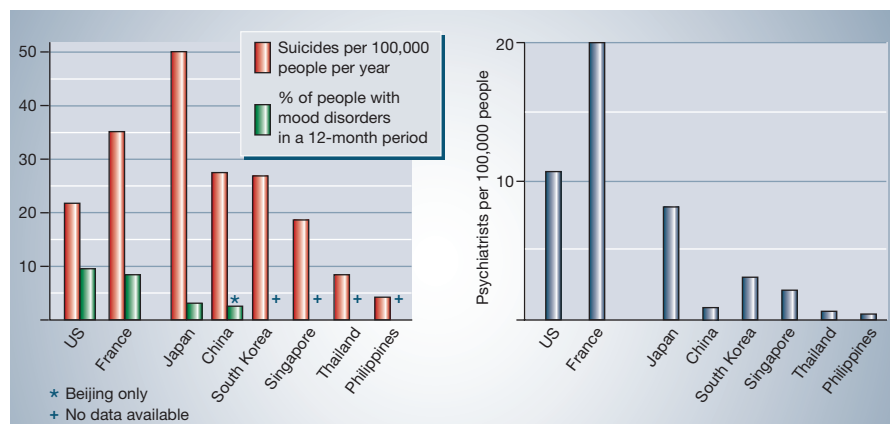
There are also marked differences in provision for psychiatric care from nation to nation. Impoverished Laos, for instance, has only two psychiatrists in the entire country. One general trend, however, is that East Asian nations have tended to hospitalize patients suffering from severe mental illness, rather than providing care in the community — which in the West has been shown to yield better results⁸. And psychiatry has traditionally been seen as a lowly specialty, marginalized in medical training.

Given their problems with mental health, some Asian nations are now trying to improve the status of psychiatry, and to move towards Western-style community care. But it won't be an easy transition. "An urgent priority is the development of an effective and equitable health system," says Harry Minas of the University of Melbourne in Australia, who, together with collaborators at Harvard University, runs a programme to train psychiatrists from across East Asia.

Drug targets

Now in its third year, the programme enrolls a dozen or so psychiatrists each year, engaging them for 12 months in a regional network involved in healthcare reform stretching from Indonesia to South Korea. Participants also spend four weeks in Melbourne attending training seminars and visiting local mental-healthcare services. One alumnus has already been inspired by the experience to open Vietnam's first community mental-health centres in Hanoi.

But for the time being, most of the overstretched psychiatrists across the region must rely heavily on the use of antidepressant drugs — in many cases drugs that are considered obsolete in the United States and Europe. Keenly aware of a huge potential market for their newer products, multinational pharmaceutical companies are now trying to raise awareness of depression,



SOURCE: WHO/REF: 3/WWW.CVDINFOBASE.CAMH-ATLAS

anxiety disorders and other mental illnesses across Asia. But tailoring antidepressants developed for Western populations to Asian patients will require research into genetics.

"In general, most Asian patients respond to lower doses of psychotropic drugs than Caucasians," says Keh-Ming Lin, who heads the Division of Mental Health and Drug Abuse Research at the National Health Research Institutes in Taipei, Taiwan. In some instances, the dose can be halved.

Lin was among the first psychiatrists to recognize that drugs used to treat mental illness have different effects across ethnic boundaries. More than 15 years ago, while working at the University of California at Los Angeles, he found evidence that Asian patients metabolize haloperidol, commonly used to treat schizophrenia, more slowly than Caucasians⁹. Asian patients are also more likely to develop severe side effects when treated with antipsychotic drugs¹⁰.

Unpublished work from Chee Ng, a psychiatrist at the University of Melbourne, suggests that it's a similar story for antidepressants. He has found that patients of Chinese origin respond to lower doses of an anti-depressant called sertraline. These patients were also more likely to drop out of the study. "They may experience more side effects at regular doses and therefore be less compliant in taking the medication," Ng suggests.

These phenomena are probably due to genetic differences that alter the activity of enzymes that metabolize drugs and allow them to be removed from the body. The most striking examples are the cytochrome P450 enzymes in the liver and, in particular, an enzyme called CYP2D6. "Almost half of the drugs used in psychiatry are metabolized by CYP2D6," says Lin, adding that up to 70% of Asians harbour subtle mutations that impair the enzyme's activity.

Rather than making gross generalizations about the response of Asian patients to drugs, geneticists argue that more detailed studies are needed to look at the relationship between genetic make-up and drug response for each pharmaceutical in different populations. In these studies, it will be important to take account of environmental factors, including diet, smoking and alcohol consumption, which can affect an individual's response to a drug.

Another key variable is whether patients are supplementing their medication with herbal remedies, which can interfere with the metabolism of other drugs. In many Asian cultures, it is common for patients to try traditional medicine before turning to Western medicine. "We are not the first choice," says



Worlds apart: different methods are needed to diagnose psychiatric illness in urban and rural China.

Nor Hayati Ali, a psychiatrist at Kuala Lumpur Hospital in Malaysia. "Most patients still consult a traditional healer first."

Cultural revolution

Understanding how different populations respond to different medicines is just one challenge for researchers tackling mental illness in Asia. Different cultures are likely to respond differently to other treatments, such as psychotherapy and behavioural therapies. So psychiatrists will need to investigate how to tailor these therapies to work effectively in various Asian countries.

Researchers also need to get better estimates of the prevalence of mental disorders across the region. Epidemiological studies have so far been hampered by inconsistencies in the diagnostic criteria used.

Improving matters will require the development of diagnostic tools that are sensitive to different ways in which mental disorders can manifest themselves in different Asian cultures.

Although Asia's problems with mental illness are daunting, psychiatrists such as Phillips are inspired by the knowledge that they can make a difference. A Canadian national, Phillips first visited China in 1976 with friends when he was a recently qualified doctor working in New Zealand.

Months later, he returned to China on a fellowship with the goal of studying how government propaganda is used to promulgate public-health issues, gaining information he hoped to apply in another developing region. In the event, the Chinese authorities did not let this project go ahead, but Phillips stayed nonetheless and learnt Mandarin Chinese. Within two years he was hooked, both on Chinese culture and the feeling that he could make a difference.

After moving to the United States to train



in psychiatry, epidemiology and anthropology, Phillips settled in Beijing in 1985. In the early days, he was regarded with mistrust. "There was a suspicion that I might be a spy," he says. But today, he finds that being a foreigner has some advantages, allowing him to be more outspoken in his relations with the media than his Chinese colleagues. "It helps me get my message out," he says.

Phillips' message is getting through: the Chinese authorities now support his research. In one project, he has interviewed a sample of 1,000 people from the general population to understand the factors that underlie the stigma associated with suicide. "To change attitudes, you need to know what the attitudes are," he says. In another study, his team is following up on rural people who have attempted suicide, and offering them social support, such as trying to ameliorate family crises that can underlie their mental-health problems. The goal is to develop a model for suicide prevention in rural China.

"Suicide is the number one cause of death in young Chinese, and up until ten years ago no one was doing anything about it," says Phillips. Now numerous hotlines and prevention programmes are starting to spring up. "China is changing so rapidly," he says. "It is amazing to watch and to be part of it."

Carina Dennis is Nature's Australasian correspondent.

1. Lin, K.-M. & Cheung, F. *Psychiatr. Serv.* **50**, 774-780 (1999).
2. Kleinman, A. *Social Origins of Distress and Disease: Depression, Neurasthenia, and Pain in Modern China* (Yale Univ. Press, New Haven, Connecticut, 1986).
3. The WHO World Mental Health Survey Consortium. *J. Am. Med. Assoc.* **291**, 2581-2590 (2004).
4. Phillips, M. R., Li, X. & Zhang, Y. *Lancet* **359**, 835-840 (2002).
5. Phillips, M. R., Liu, H. Q. & Zhang, Y. P. *Cult. Med. Psychiatr.* **23**, 25-50 (1999).
6. Kleinman, A. *Cult. Med. Psychiatr.* **6**, 117-189 (1982).
7. *Report on the WHO Collaborative Study on Standardized Assessment of Depressive Disorders* (WHO, Geneva, 1983).
8. *The World Health Report 2001* (WHO, Geneva, 2001).
9. Lin, K.-M. et al. *J. Clin. Psychopharmacol.* **8**, 195-201 (1988).
10. Lin, K.-M. & Poland, R. E. in *Psychopharmacology: The Fourth Generation of Progress* (eds Bloom, F. E. & Kupfer, D. J.) 1907-1917 (Raven, New York, 1995).

European science must embrace modernization

Researchers in the United States benefit from having more money, mobility and flexibility.

Sir—As *Nature* has noted in recent News stories and Editorials, French researchers are revolting against their government's strategic policy on science, while the government criticizes the ossified structure of the research establishment (see *Nature* 428, 105 & 108; 2004). The German government has recently announced its intention of creating elite universities to match the best in the world (*Nature* 427, 271 & 477; 2004). And changes to UK research funding will probably concentrate resources in a small number of large and prestigious centres (*Nature* 428, 351; 2004).

These events indicate a major government-initiated modernization in science policy. European scientists who wish to work in an efficient system that produces top-quality work should embrace, not oppose, this new wave of modernization.

One reason for reform is the growing realization that — according to several quality and efficiency criteria, such as citation analyses — science in the United States is outperforming European science. And this gap is widening, especially when it comes to generating research of major importance (*The Economist* 369, 5–7; 2003).

One oft-cited difference between Europe and the United States is science funding, which is proportionately higher in the United States and concentrated in relatively fewer institutions.

However, as well as having more money, US science enjoys a greater diversity of independent public and private funding sources, a situation favoured by tax regulations and greater institutional autonomy (for example, in private research institutes).

There are also important differences

in the structure of scientific careers, such as the greater mobility of US researchers, including senior scientists, and the greater ease of hiring and funding foreign researchers and PhD students.

Together, these factors largely prevent the ossification of research institutions, and fuel productivity and innovation.

If European governments wish to match the results of US science, policy reforms will need not merely to address the funding inequality but also to create career and funding structures that generate increased competition and differential rewards among both scientists and research institutes.

Peter Andras*, **Bruce G. Charlton†**

**School of Computing Science,
University of Newcastle,*

Newcastle upon Tyne NE1 7RU, UK

*†School of Biology, University of Newcastle,
Newcastle upon Tyne NE2 4HH, UK*

Reliable regional climate model not yet on horizon

Sir—We agree with the overall thrust of your News story “Modellers deplore ‘short-termism’ on climate” (*Nature* 428, 593; 2004), with one exception. The US National Assessment of the Potential Consequences of Climate Variability and Change (USNA) — in which we were involved — did not attempt to provide regional or even national predictions of climate change, as was implied in the article. Nor, in our opinion, did it oversell the capacity of regional climate models to provide useful information.

The USNA report was explicit about its use of a range of approaches (mostly using projections from global climate models) to generate plausible scenarios of future climatic conditions that could then be used to explore the potential consequences of these scenarios for the environment, natural resources and people.

When the USNA performed regional analyses of global climate model projections, for specific areas of the United States, it did so in a way that cancelled out much of the systematic bias of the global models. This approach allowed the USNA to explore, for example, the responses of vegetation cover, agricultural production and water resources to the general character of the climate change expected during the twenty-first century.

Part of what the USNA learned and presented was a very clear description of

the confidence the authors felt could be assigned to their conclusions. In most cases, the lack of true predictability of the climate results meant that the USNA presented many caveats to its findings.

Even a cursory reading of the USNA findings (www.usgcrp.gov/usgcrp/nacc) makes clear that they are not dependent on the regional details of the projected changes in climate.

For example, global warming is very likely to lead to reduced spring snowpack in the western mountains of the United States, reducing water resources; greater evaporation during the summer in the central states, lowering river and lake levels; a higher heat index that would endanger health in the humid eastern and southeastern regions; significant shifts in the landscape; rising sea level that will affect several low-lying regions, and so on.

We strongly agree that much more reliable regional climate simulations and analyses are needed. However, at present, as the News story makes clear, such simulations are more aspiration than reality.

Michael MacCracken*, **Joel Smith†**,
Anthony C. Janetos‡

**Climate Institute, 1785 Massachusetts Avenue,
NW, Washington DC 20036, USA*

*†Stratus Consulting, PO Box 4059, Boulder,
Colorado 80306-4059, USA*

*‡The H. John Heinz III Center for Science,
Economics and the Environment,
1001 Pennsylvania Avenue, NW, Suite 735 South,
Washington DC 20004, USA*

Fame and popularity are no bar to Royal Society

Sir—Contrary to your News in Brief story, “Popularizer Greenfield is blackballed by peers” (*Nature* 429, 9; 2004), Susan Greenfield has not been “blackballed” by the Royal Society, nor have we “decided not to admit” her. Your report neglected to point out that candidates are eligible for election as a fellow for up to seven years after nomination, in the first instance, and they are considered afresh each year by new panels of distinguished scientists.

Professor Greenfield is a candidate for the fellowship of the Royal Society on the basis of her substantial contribution to science. Like all the other 535 candidates this year, she was nominated by current fellows.

It is also untrue that we “declined to go into detail about the nomination process”. The details of the nomination and election process are published in the yearbook of the Royal Society and on our website (www.royalsoc.ac.uk). For obvious reasons, the discussion by the panels of individual candidates remains confidential. Indeed, the identities of all candidates should remain confidential and it is deplorable that there has been a breach in Professor Greenfield's case.

John Enderby, **David Read**

*The Royal Society,
6–9 Carlton House Terrace,
London SW1Y 5AG, UK*

Beyond artificial intelligence

Does a program help us to make sense of the world?

What is Thought?

by Eric B. Baum

Bradford (MIT Press): 2004. 495 pp.

\$40, £25.95

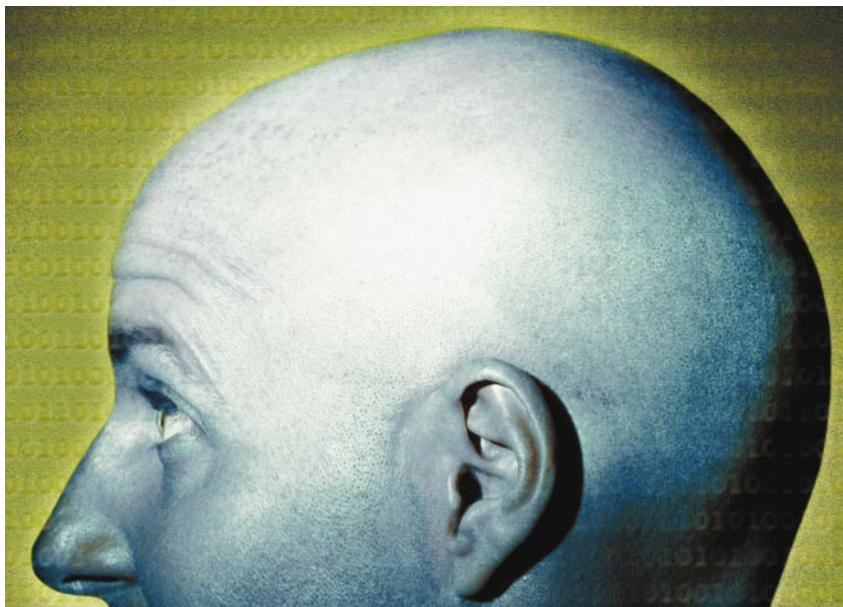
Igor Aleksander

"What is thought?" is not a new question. For Aristotle, thought was what the soul does, and for Descartes it was the unequivocal evidence of one's existence. For Eric Baum, a US expert in machine learning, it is a computer program. This is not a superficial assertion: Baum pursues the idea with elegance, clarity and considerable persuasion. His choice of book title was inspired by Edwin Schrödinger's book *What is Life?*, which was published before the discovery of DNA. Schrödinger felt that the answer to "What is life?" must lie within physics and chemistry, despite their apparent failure to deliver this at the time. Similarly, Baum argues that an answer to "What is thought?" must lie within computer science, despite its failure to produce convincing results so far under the guise of artificial intelligence.

It is important not to treat the idea that thought is a program in too superficial a fashion. Popular texts often include explanations such as "brain is like hardware and mind is like software". Baum intends a level of sophistication far above this. The program to which he refers is one that extracts meaning from complex data. Thought for him is the process that 'understands' the complexities of the world. So a thought program is one that detects sensory information as a compact code.

For example, when walking down the road we may see a dog coming towards us. Is it friendly or not? We will need to take different actions depending on the answer. What thoughts are taking place in our heads and where do they come from? They involve a heavy dose of recalled experience and even just-learned information. We might once have been bitten by a pit-bull terrier, or just know from pictures or descriptions by others that if the dog is unmuzzled and untethered then it might be best to seek refuge behind a door, for example.

All the information we need to make our decision is there in the sensory world, but its meaning needs to be decoded from myriad other data that are not relevant to the key issue, danger in this case. There may be people, cars, other dogs, leaves carried on the wind and so on. But our thought decodes 'danger' and plans a reaction to it. Calling this thinking activity a 'decoding and planning program' requires several caveats. First, it is unlikely that a human could have written



Thinking ahead: is thought a program that decodes sensory information so we can make decisions?

a program that could lead to a lifetime of thought. On the other hand, evoking an Almighty programmer is a recourse to mysticism that would leave us befuddled rather than enlightened. Baum's central point is that it is quite possible for programs to evolve, adapt and learn, making them more powerful than anything that a programmer can concoct.

A second caveat is that Baum does not suggest that someone is going to create an evolutionary, adaptive, learning program, put it into a robot and create a thinking object indistinguishable from a human being. Rather, Baum's argument is that a good (if not the best) way of understanding human thought is to analyse it as if it were a program. Artificial intelligence, in the past, was the product of programmers writing smart programs that do clever things, such as beating Garry Kasparov at chess. Baum goes beyond this by presenting clear explanations of what it is to extract the compactly coded information in the world using the simplest possible program; what it is for such a program to come into being through a process of evolution and adaptation; and what it is for a program to learn both over several generations and during daily life. Baum calls this 'Occam's razor' programming, stressing that the simplest program model is likely to have the best powers of explanation.

The third caveat is my own. One must understand that Baum's explanation is a metaphorical or functional one — it is not

a material model of the brain. Even though some of his computational models are 'neural networks' (a technology loosely inspired by the operation of brain cells), Baum does not suggest that a careful analysis of the brain will reveal that it is structured like a general-purpose computer on which the meaning-extraction programs run. His point is that meaning computation is a powerful explanatory metaphor for what the brain, its complex neurons, its chemical transmitters, its muscular engagement with the world and its highly adapted and impenetrable architecture do in an abstract sense. Although Baum does attempt to answer questions about what kind of program is needed for awareness, consciousness and will, he does so at a metaphorical level and does not address what the material brain does that makes it appear to a computer theoretician to be running these programs.

An important part of the book is devoted to the considerable flak that artificial intelligence scientists attracted in the heyday of this topic. Baum gives a reasoned response to John Searle's claim that no program can 'understand' the world, and to Roger Penrose's contention that conscious insight lies outside the logic that can be achieved by computation. The essence of the response is that these objections were made before ideas such as evolution, adaptation and compact decoding had become part of the idea of computation. Baum joins in the criticism of what is now called GOFAI, or 'good old-fashioned artificial intelligence', so

D. TEEL/CORBIS

this is a splendid book for discovering what is new. It will enthral some computer scientists and provoke some philosophers. And it should engage general readers who wish to enjoy a clear, understandable description of many advanced principles of computer science. ■

Igor Aleksander is emeritus professor and Leverhulme fellow in intelligent systems and networks at Imperial College, London SW7 2BT, UK. He is writing *Five Steps to Being Conscious: the Inner Life of People, Animals and Machines*.

Split decision

The Fly in the Cathedral: How a Small Group of Cambridge Scientists Won the Race to Split the Atom

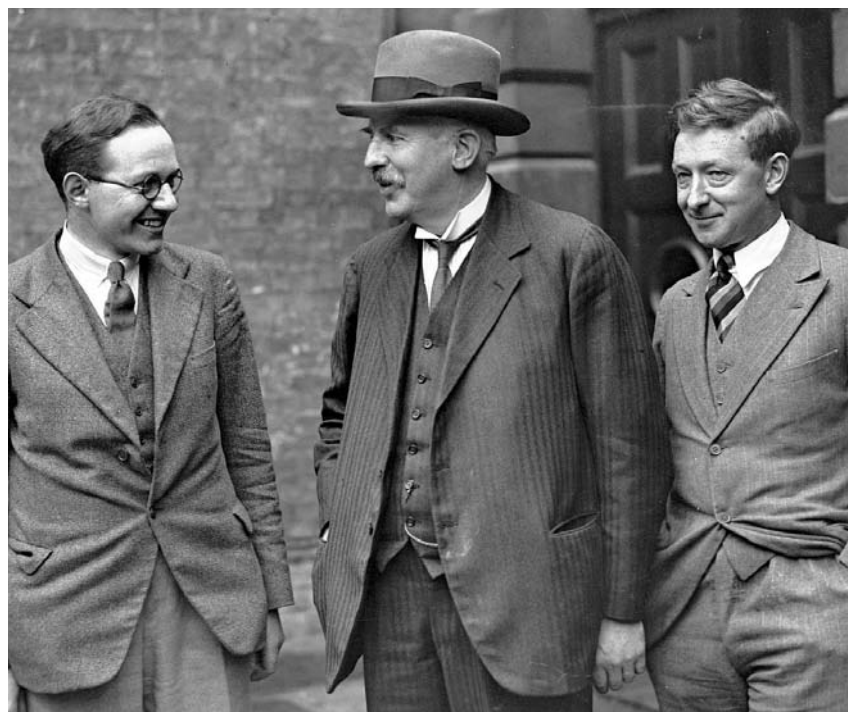
by Brain Cathcart
Viking: 2004. 320 pp. £14.99

Frank Close

If an atom were expanded to the size of a cathedral, its nucleus would be the size of a fly. Ernest Rutherford deduced the existence of this atomic 'fly' some 90 years ago, but the challenge of illuminating its mysteries and determining its make-up seemed to be beyond reach. Then, after years of effort, two young researchers, John Cockcroft and Ernest Walton, under Rutherford's guidance, built their eponymous accelerator in a basement at the Cavendish Laboratory in Cambridge, UK. In 1932 the media heralded their triumph with a phrase that has rung through the ages: "The atom split". The age of accelerators and big science had arrived. *The Fly in the Cathedral* tells the story of how this came to be.

Books about theoretical physics abound; those about experiments are thinner on the ground. *The Fly in the Cathedral* is indeed a rarity in focusing on how a classic experiment came about. The book's strength is that it reveals the nature of science — in particular how inspiration is partner to vast amounts of perspiration, commitment and dedication.

Rutherford discovered the existence of the atomic nucleus by scattering alpha-particles from it. No one knew how a positively charged alpha-particle managed to survive inside the atom along with all the other positive charges. But they knew that some were ejected "as if hurled by a catapult" at tremendous speed, thanks to repulsive electric forces. It was by scattering these natural bullets from atoms that Rutherford came to discover the atomic nucleus. To understand how the nucleus was constructed, it would be necessary for beams of alpha-particles to penetrate it, but first they would have to be given even higher energies to overcome the huge electrical barrier. An



Ernest Rutherford (centre) helped Ernest Walton (left) and John Cockcroft to split the atom.

accelerator powered with millions of volts was needed. This was what Cockcroft and Walton set out to make.

Quantum mechanics came to the rescue. George Gamow showed that alpha-particles could be trapped within a huge potential barrier inside the nucleus and that, according to quantum mechanics, there was a chance that the alpha-particles could 'tunnel' through it and escape. His beautiful exposition is now a standard example in undergraduate quantum-mechanics courses. A good idea is rarely yours alone, though, and Gamow was beaten into print by Edward Condon and R. W. Gurney, even though their letter was dated one day after his. His fortune was that he came to the attention of Rutherford, which in turn led to his move to Cambridge.

Gamow gained his own priority, though, with a seminal paper in *Nature* in which he turned the original question on its head. Instead of discussing how particles manage to escape from the nucleus, he looked at how they might get in. Quantum mechanics enables alpha-particles to penetrate the potential Coulomb barrier without the need for them to be accelerated to huge energies. Cockcroft realized that if protons were used instead of alpha-particles, things would be even better. Instead of millions of volts, as they had originally thought, as little as 300,000 volts might be enough. This opened the way for the experimental study of nuclear physics.

In Cambridge, Cockcroft and Walton were hard at work building their device. In the United States, Ernest Lawrence was

building the first cyclotron. This promised far greater energies than Cockcroft and Walton's device, but Lawrence's claims were based more on hope than realization. The pace of Brian Cathcart's narrative slows somewhat when describing the details of Cockcroft and Walton's efforts, but the slightly ponderous tone at this point is a metaphor for what actually happened: Cockcroft and Walton seemed obsessed with building and improving their machine until Rutherford one day insisted that they actually put it to work.

Suddenly the story comes alive again. Gamow's estimates, which everyone was treating with caution and aiming to beat, turned out to be conservative. Protons would invade a nucleus at relatively moderate energies easily within the reach of Cockcroft and Walton's device. It transpired that they could have achieved their results with the prototype accelerator they built some three years earlier. The rest, as they say, is history.

In addition to the history, it is unusual to see the portraits, both literal and visual, of the protagonists and appreciate their commitment to seeing the task through. For all the hyped speculation about superstrings, higher dimensions and theoretical science that borders fact and fiction, it is good to see books such as *The Fly in the Cathedral*, which remind us that it is ultimately experiment that decides. And that creating tools to expand our senses can provide a fascinating and dramatic story. ■

Frank Close is at Exeter College and the Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford University, Oxford OX1 3NP, UK.

To see and not to see

Sight Unseen: An Exploration of Conscious and Unconscious Vision by Melvyn Goodale & David Milner
Oxford University Press: 2003. 160 pp. £25, \$49.50

Manfred Fahle

This book was published at the wrong time, I think. Judge for yourself. It tells the story of two patients, D.F. and Ruth Vickers, and their astonishing perceptual deficits. At the age of 35, D.F. suffered from carbon monoxide poisoning caused by a defective Italian gas heater. (By the way, did you know that carbon monoxide poisoning is the number one cause of unintentional poisoning deaths in the world?) As a consequence, D.F. is unable to visually recognize the forms and shapes of objects — shapes seem to “run into each other”. This disorder is called apperceptive agnosia, or form agnosia. Nevertheless, D.F. appears normal when walking, even negotiating mountain paths, and so must have retained the ability to use vision to navigate through the world.

Ruth Vickers, on the other hand, suffered from bilateral lesions of another part of the cortex, has unimpaired perception of objects and can name objects presented to her visually. Ruth's motor system is also intact. However, she struggles to grasp objects that she can clearly see — a deficit called optic ataxia. The dissociation of defects in the perception of objects and in handling visually presented objects indicates that two relatively independent subsystems exist in the human brain: one specialized for identifying visually presented objects, the other for visually guiding actions.

Given the the authors' clear and precise language and their stated aim to write an accessible book (which they achieve), this volume is a perfect Christmas present for anyone even remotely interested in the brain. And that's why it was published at the wrong time — you will have forgotten this review by Christmas.

If the description so far resembles one for another book by Oliver Sacks, it should not. *Sight Unseen* is not just a book for readers of popular science, demonstrating how much can be learned about brain function from patient studies; even specialists in neuroscience and neuropsychology could learn something. For example, separate text boxes review specific topics; one lists about ten different brain regions to which retinal fibres directly project. The text, too, contains a lot about cortical specialization that most neurobiologists probably do not know, making it easy and rewarding reading for specialists who are willing to accept that

Exhibition

Retreat of the ice giants

They are a playground for summer skiers, a challenge for mountaineers, and their dazzling whiteness under a deep-blue sky is of other-worldly beauty. But glaciers are also crucial for Alpine water balance — and they are a unique indicator of global warming.

Since the Industrial Revolution, glaciers in the Alps have decreased by more than one-third in size and by a half in volume. This dramatic retreat of supposedly eternal ice has now been visualized by a couple of German photographers who are also environmental scientists.

Wolfgang Zängl and Sylvia Hamberger collected hundreds of historic postcards of glaciers in Germany, Austria, Switzerland, Italy and France. They carefully determined the month and time of day when each picture was photographed or painted, and identified the artists' viewing points. Then they took similar



shots of the landscapes as they appear today. The pictures shown here are of the Rhône glacier in Vallais, Switzerland.

The results are now on show in the Alpine Museum in Munich, Germany. The exhibition

“Glaciers in the Greenhouse”, which runs until 16 January 2005, also provides some numbers and diagrams to illustrate the problem of global warming in the twentieth century. But its visual centrepiece, the pairs of old and new glacier photographs, depict much more shockingly what climate change is doing to the Alps. **Q.S.**

some problems cannot be thoroughly discussed in a book of this sort.

The main purpose of the book is not to describe the surprising symptoms that arise in patients after damage to the brain, but to produce a general picture of brain function. This bigger picture starts with the generally accepted insight that our nervous system has to make sense of the world. Hence, the brain cannot simply reproduce the world but must interpret it. (Who would be there to look at the television screen inside our head?) The authors argue that part of this interpretation takes place at a subconscious level. They claim throughout the book that there is a system for the conscious analysis of visual objects, and a faster one that acts outside consciousness to guide our actions.

This separation, derived from patient studies, makes sense from a computational point of view. To identify objects reliably, the visual system must abstract them from such accidental and varying factors as their position, distance and orientation — a rose is a rose irrespective of its position and distance. When I intend to pick up the rose, on the other hand, these factors become very important. So it makes sense to use one system that generalizes over space to identify objects and another, faster one that only cares about their position and orientation, without worrying too much about their identity or involving the conscious ‘I’.

Combining both tasks — identification and localization — in a single system would be asking too much.

Don't start feeling smug, though, simply because you have two visual systems rather than just one. In 1973, D. J. Ingle published a paper entitled “Two visual systems in the frog” (*Science* **181**, 1053–1055), and then L. G. Ungerleider and M. Mishkin reported that the same is true for monkeys, in *The Analysis of Visual Behaviour* (MIT Press, 1982) edited by Ingle, R. W. Mansfield and Goodale.

In conclusion, Goodale and Milner emphasize that much of what goes on in our brains, and even in our cortices, escapes our conscious ‘I’; partly because of the separation of the visual systems for perception and action. As a result, some patients can see and not see — being unable to visually guide their actions — whereas others can navigate visually without seeing what objects are there.

The book illustrates the enormous amount of knowledge to be gained from analysing deficits of specific stroke patients. It closes by stating: “Studying the way the brain reorganizes itself in response to severe damage presents one the most important challenges to neuroscience in the twenty-first century.” How true. ■

Manfred Fahle is at the Center for Cognitive Sciences, Bremen University, 28211 Bremen, Germany.

The ties that bind

Attachment: the nature of the bonds between humans are becoming accessible to scientific investigation.

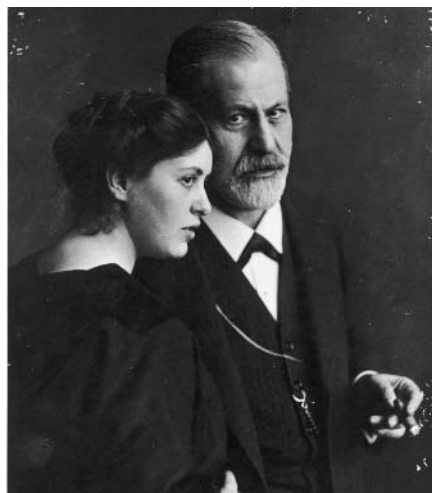
Melvin Konner

Attachment is the name we give to bonds between people. It has been central to song and story since the dawn of human time, but has only recently become a subject of scientific study. Sigmund Freud had much to say about how the mind handles it, but conceded that “our provisional ideas in psychology will someday be based on an organic substructure”. Today, we have glimmerings of that substructure.

John Bowlby emphasized the most basic attachment, that of an infant toward its primary caregiver. Bowlby’s model of attachment was informed by evolution — eons of selection had pressed mothers and infants into each others’ arms. The notion shared by Freud and B. F. Skinner (otherwise sworn enemies) — that infants became attached through reinforcement of their hunger drive — had failed decisively. Harry Harlow demonstrated that “love in infant monkeys” transcended such simplicities when a wire-mesh surrogate mother supplying delicious milk lost out in the battle for infant attachment to another inanimate surrogate providing only warmth and contact comfort. From this and other observations, Bowlby reasoned that attachment was something built into infants and was programmed to unfold on a predetermined schedule. Anthropological evidence supports the general model. In all cultures, attachment behaviours — such as turning and clinging to the primary caregiver in distress and privileging that person by preferentially quieting the distress — becomes very strong in the second half-year of life.

It is probably not a coincidence that in the brain, major pathways of the limbic system become coated with myelin during this phase of infancy. This improves the function of the subcortical circuits that process emotion and their connections to the frontal and cingulate cortex. Although there is no direct evidence, it is reasonable to hypothesize that this facilitates the infant’s side of the bond.

For the other half of the relationship, oxytocin is vital in many non-human mammals. This peptide hormone, also involved in milk let-down and uterine contractions, causes mothers to retrieve and respond normally to infants. Oxytocin knockout mice develop a strange social amnesia. And vole species with strong maternal behaviour have a different and denser distribution of oxytocin receptors in the brain than closely related species where maternal behaviour is weaker.



Fatherly feelings: Freud and his daughter Sophie.

Sue Carter has shown that this brain pattern is also associated with other forms of affiliative behaviour, not just in the maternal realm.

But getting males attached to infants — or their mates — involves a different hormone, as Thomas Insel has shown. Vasopressin is well known to physiologists as being crucial for fluid balance. Like oxytocin, vasopressin is a brain neurotransmitter. Both evolved by one amino-acid substitution from the hormone vasotocin, which in reptiles such as the sea turtle plays a role in nest-building and egg-laying. The much more complicated parental behaviour of mammals required a more refined system.

Male prairie voles are paternal and pair-bonding, whereas montane voles have multiple-mating males that leave the care of the young to the females. The dedication of the prairie voles is due to vasopressin receptors that are distributed strategically throughout the male brain — especially the V1a receptor. On page 754 of this issue, Miranda Lim and colleagues in Larry Young’s laboratory have built on Carter and Insel’s work. Meadow voles are normally promiscuous, but one gene can change all that. Introducing the prairie vole V1a in a viral vector delivered to the fore-brain makes male meadow voles mate for life. As in prairie voles, these genetically modified meadow voles have dopaminergic reward circuits activated by vasopressin — this could mean they not only commit, but like it. This idea is discussed in more detail by Evan Balaban on page 711. We do not yet know if a similar system helps explain male attachment in non-human primates, much less humans, but a medicine that might someday be offered to certain men is an interesting prospect.

We are a long way from a commitment pill, but perhaps closer to a neurology of romance. Recent studies have identified the anterior cingulate cortex, the insula and the caudate nucleus as unusually active in head-over-heels subjects, who also have a low density of serotonin transporters in blood platelets. (Perhaps unsurprisingly, measures for obsessive-compulsive patients looked a lot like the besotted lovers serotonin-wise.)

There are many questions. What happens if romantic fancy becomes long-term commitment? Do oxytocin and vasopressin, those neurochemical workhorses of prairie vole bonding, take over in humans too? Parents and infants may fall in love with each other, but are they using the same circuits as grown-up star-crossed lovers? When people ‘learn to love each other’ after arranged marriages — most since the world began have been in that category — do they look, neurologically, like the quietly loving bond that once began in a burst of romantic delirium?

Some questions are more pragmatic. New large-scale studies from the United States suggest that lower-quality day care may alter infant attachment and cortisol patterns. Does this effect bode ill for the person the infant will become, or is this just a different path on the way to adapted adulthood? At the other end of the spectrum, severely deprived infants — Romanian orphans, for example — sometimes have a syndrome called ‘reactive attachment disorder’, which entails abnormalities in later relationships. Perhaps brain imaging and neuropharmacology will yield treatments for such disorders, and may even help with autism-spectrum syndromes. However amorphous attachment may seem to a physicist, it is one of the most important determinants of human well-being, and we would do well to bring it into scientific focus. ■

Melvin Konner is in the Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA, and is the author of *The Tangled Wing: Biological Constraints on the Human Spirit* (revised edn).

FURTHER READING

- Bowlby, J. *Attachment and Loss* (3 vols) (Hogarth Press, London, 1969–1977).
- Cassidy, J. & Shaver, P. R. (eds) *Handbook of Attachment: Theory, Research, and Clinical Applications* (Guilford, New York, 1999).
- Konner, M. in *Hunter-Gatherer Childhood* (eds Hewlett, B. & Lamb, M.) (Aldine, New York, in the press).
- Insel, T. R. *Rev. Gen. Psych.* **4**, 176–185 (2000).
- Fisher, H. E., Aron, A., Mashek, D., Li, H. & Brown, L. L. *Arch. Sexual Behav.* **31**, 413–419 (2002).
- Bartels, A. & Zeki, S. *NeuroImage* **21**, 1155–1166 (2004).

On Earth, as it is on Mars?

David C. Catling

The small spheres of haematite, nicknamed 'blueberries', that litter the Mars landing site of NASA's rover Opportunity might have an analogue on Earth, formed from groundwater in southern Utah.

In Soviet times, Russians would joke, "Is there life on Mars? No, not there either". Indeed, by the late 1970s, NASA's Viking mission had revealed a cold, dry planet, hostile to life¹. Then, in the 1990s, came the confluence of two strands of science: a heated debate about possible traces of ancient life in a martian meteorite² and, in biology, increasing evidence that microbes can thrive in extreme environments. The idea of life on Mars re-emerged. The question of where to look for signs of past or present life on Mars has been guided by the principle that all life, as we know it, requires liquid water. Consequently, when NASA launched two rovers to Mars last year, one of the rovers, named Opportunity, was targeted at Terra Meridiani, a region where orbital reconnaissance had indicated abundant crystalline haematite (Fe_2O_3), a mineral whose precursors usually form in water³⁻⁶.

On 25 January 2004, Opportunity landed in a 22-m-diameter crater in Meridiani. There the rover's camera stared across dark basaltic sand towards an outcrop, 0.5 m tall and 12 m long, of light-toned sedimentary rock near the crater rim⁷. Surprisingly, the haematite was found to be in the form of spheres, typically a few millimetres in diameter and embedded in rock outcrops like "blueberries in a muffin" (in the words of Steve Squyres, the lead rover scientist). But as

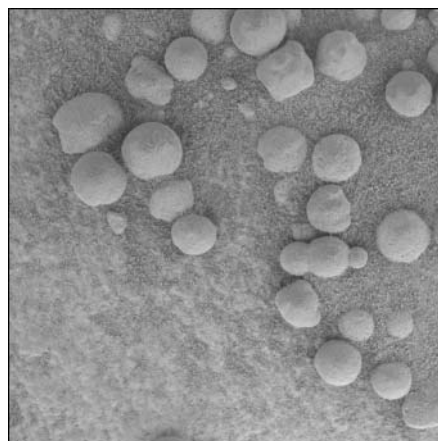


Figure 1 Earth or Mars? The NASA rover Opportunity found blueberry-size concretions of haematite on the martian surface (left), which Chan *et al.*⁹ point out are strikingly similar to these larger haematite 'marbles' found in Utah (right).

the outcrops crumble away through wind erosion, the haematite spheres, which are more resistant to erosion, are falling to the ground. Haematite spheres or their broken fragments litter the plains of Terra Meridiani and represent leftovers of a large area of rock that has vanished in the wind.

NASA's rover team has interpreted the haematite spheres as concretions, which are hard, compact mineral accumulations of different composition from their surrounding rock. Concretions form from water

that carries dissolved minerals through soft sediments or porous rock: minerals precipitate around a nucleus in layers that can incorporate or replace surrounding sediment. On Earth, nucleation is sometimes attributed to microbial processes⁸. Concretions occur in a variety of terrestrial rocks, but is there anything like the martian blueberries?

On page 731 of this issue, Chan *et al.*⁹ report on some remarkable sandstone rocks in southern Utah that are filled with

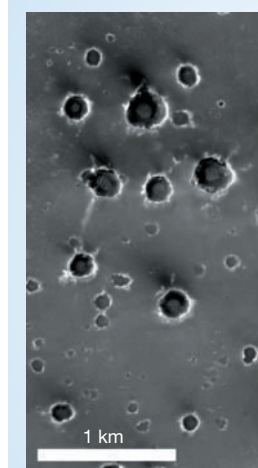
Box 1 Rock and fluid chemistry

Craters in the haematite-rich Terra Meridiani on Mars are edged with peculiar bright rings (left image). These rings can now be interpreted as outcrops of light-toned sedimentary rock, resulting from asteroid or comet impacts that created the craters by punching into underlying light-toned rock. Over geological time, dark basaltic sand and haematite granules have filled the crater interiors, and wind has degraded the crater rims.

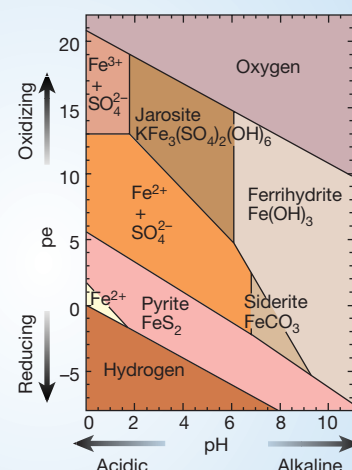
The light-toned outcrops contain haematite concretions and sulphates, including jarosite. Jarosite must have formed in an acidic and oxidizing environment, as

indicated in a pe–pH diagram (right). Just as pH is defined as the logarithmic activity of aqueous hydrogen ions ($\text{pH} = -\log_{10}(a_{\text{H}^+})$) to measure acidity or alkalinity, pe is defined analogously as the logarithmic activity of electrons ($\text{pe} = -\log_{10}(a_e)$) to indicate potential electron transfer — that is, oxidation or reduction. (The diagram is drawn for 25 °C, CO_2 at a pressure of 0.01 bar, and assumed activities of dissolved iron, sulphur and potassium of 10^{-4} , 10^{-2} and 10^{-4} , respectively. Activity is concentration corrected for non-ideal behaviour.)

Acidic water can transport iron as $\text{Fe}^{2+}(\text{aq})$, which commonly



precipitates in more alkaline and oxidizing conditions as ferrihydrite, $\text{Fe}(\text{OH})_3$. Over time and with warm temperatures, ferrihydrite



dehydrates to goethite (FeOOH) and then haematite (Fe_2O_3), displacing the jarosite stability field in the pe–pH diagram.

D.C.C.

haematite concretions. As on Mars, the weathering of these rocks causes the concretions to tumble to the ground and spread across the surrounding plains. There the concretions become smooth and shiny from wind abrasion, and their diameters range from millimetres to centimetres. The loose Utah concretions roll like marbles into depressions, forming 'puddles', just like their martian counterparts (Fig. 1).

The host rock for the Utah concretions is Navajo Sandstone. This rock derives from a vast desert of sand dunes that was deposited by trade winds in the Jurassic, when Utah was situated in the tropics. Microscopic iron-oxide films coat the sand grains and impart vivid red and pink colours to the sandstone. Fine iron oxides are reddish, unlike the brownish-black colour of densely cemented iron oxide in haematite concretions. Voids between Navajo Sandstone grains, which were sorted to uniform size by ancient winds, render the sandstone porous like a sponge. Fluids that are reducing, acidic, hot, or some combination of these, can mobilize iron. In certain areas, Navajo Sandstone has been bleached white by groundwater that dissolved the fine iron oxides. Where iron-rich fluids encountered more oxidizing or alkaline microenvironments, iron precipitated and cemented into concretions. But because iron is poorly soluble, large amounts of groundwater were needed — Chan *et al.*⁹ estimate 100 kg of water per gram of iron.

How good an analogue are the Utah concretions to those on Mars? There are important similarities, which include the surface accumulations, iron-rich fluids, porous host rock and significant volumes of fluid. But Chan *et al.* also note key differences. Opportunity's spectrometers and camera indicate that the martian haematite is pure, crystalline and grey. In contrast, the Utah concretions are mostly quartz, with brownish-black haematite cement as a secondary component. Consequently, the infrared spectrum that alerted us to haematite on Mars is unlikely to match a quartz-dominated spectrum from the Utah concretions.

Fine-grained Navajo Sandstone is comparable to the fine texture of the sedimentary rocks in Terra Meridiani. But the martian rocks are compositionally quite different and contain up to 40% sulphates, suggesting that they were deposited in an evaporating brine⁷. NASA's rover team has reported the presence of jarosite, $\text{KFe}_3(\text{SO}_4)_2(\text{OH})_6$, which, from thermodynamic considerations, implies an acidic environment (see Box 1). Thus, on Mars it was probably acidic water, rather than the reducing, hydrocarbon-rich fluids inferred for Utah, that mobilized iron. Acidity might have derived from dissolved carbon dioxide, given Mars's CO_2 -dominated atmosphere, but such a fluid would leave behind tell-tale carbonate minerals. The absence of carbonates and

the conspicuous abundance of hydrous sulphates imply that the solution probably contained sulphuric acid.

The Hopi Indians have a legend that 'moqui', or spirits of their ancestors, played games of marbles with the haematite concretions in the American southwest. Although anthropologists discourage use of the word 'moqui', to be respectful to Native Americans, New Age gem collectors sell concretions as 'moqui marbles' and claim that they are endowed with metaphysical powers. New Agers are at least partly correct in their supposition that the haematite marbles can provide answers to the questions that you might ask of them. Given the similarities⁹ between the marbles in Utah and on Mars, additional scientific scrutiny of the Utah concretions and how they form will probably shed further light on the similar phenomenon on Mars.

Indeed, if we could bring back concretions from Mars, much could be learned. On Earth, concretions can be valuable indicators of sediment burial history, physico-chemical

conditions, palaeobiology and even palaeoclimate¹⁰. Moreover, the detailed geochemistry of the sulphate-rich martian rocks would probably prove even more revealing. So perhaps NASA's proposed sample-return mission for 2013 should shoot for Terra Meridiani, and follow the wheel tracks of Opportunity. ■

David C. Catling is in the Astrobiology Program and the Department of Atmospheric Sciences, Box 351640, University of Washington, Seattle, Washington 98195, USA.

e-mail: dcatling@u.washington.edu

1. Klein, H. P. *J. Geophys. Res.* **103**, 28463–28466 (1998).
2. McKay, D. S. *et al. Science* **273**, 924–930 (1996).
3. Christensen, P. R., Morris, R. V., Lane, M. D., Bandfield, J. L. & Malin, M. C. *J. Geophys. Res.* **106**, 23873–23885 (2001).
4. Lane, M. D., Morris, R. V., Mertzman, S. A. & Christensen, P. R. *J. Geophys. Res.* **107**, doi:10.1029/2001JE001832 (2002).
5. Catling, D. C. & Moore, J. M. *Icarus* **165**, 277–300 (2003).
6. Baldrige, A. M. & Calvin, W. M. *J. Geophys. Res.* **109**, doi:10.1029/2003JE002066 (2004).
7. Squyres, S. W. & Athena Science Team *Int. J. Astrobiol.* **3**, Suppl. S1, doi:10.1017/S14735500404001648 (2004).
8. Coleman, M. L. *Mar. Geol.* **113**, 127–140 (1993).
9. Chan, M. A., Beitler, B., Parry, W. T., Ormó, J. & Komatsu, G. *Nature* **429**, 731–734 (2004).
10. Sellés-Martínez, J. *Earth Sci. Rev.* **41**, 177–210 (1996).

Molecular biology

The loader of the rings

Michael A. Trakselis and Stephen D. Bell

Among the numerous molecular machines involved in the process of DNA replication are the ring-shaped sliding clamp and the clamp loader. Intriguing structural details of their interaction are now revealed.

The double-helical structure of DNA is an icon of our time, appearing almost daily as a backdrop to news stories about medical advances and in myriad sci-fi movies. The molecule itself consists of two inter-wound strands, the backbones of which are composed of alternating phosphate and sugar groups. Extending from the backbone into the heart of the double helix are the bases — guanine, adenine, thymine and cytosine — whose order encodes the information stored by DNA. The bases on one strand must pair precisely with those on the opposite strand, adenine with thymine and guanine with cytosine¹, and this provides a simple and elegant mechanism for copying this genetic material. The double helix simply unwinds and unzips, with both strands then being used as templates for the enzyme DNA polymerase, together with a large assembly of accessory factors, to make two daughter DNA molecules. Papers published on page 724 of this issue² and in *Nature Structural and Molecular Biology*³ now add to our understanding of a crucial molecular contributor to this process.

Because of a chemical asymmetry in the arrangement of DNA, one daughter strand, termed the leading strand, is synthesized continuously during DNA replication. The

other, the lagging strand, is made in short segments (called Okazaki fragments) that are later joined together. Central to this process is the sliding clamp, a ring-shaped, multi-subunit molecule that encircles the DNA and binds to the DNA polymerase. As the name suggests, the sliding clamp can slide along DNA, and so provides a mechanism for tethering the DNA polymerase to the template. On the leading strand this is extremely important, because the polymerase might have to continue synthesizing DNA for more than a million bases. But the clamp is also important on the lagging strand, because several of the factors involved in processing the Okazaki fragments use it as a scaffold for their assembly⁴.

The ring-shaped nature of the sliding clamp presents a simple topological problem: how is DNA introduced into the hole in the clamp's centre? This is the job of a complex molecular engine termed the clamp loader, which must somehow open the sliding clamp, pass DNA into the ring and then reseal it. These requirements are a common theme in several DNA-based processes. Higher organisms, for instance, also possess a DNA-repair-specific sliding clamp and clamp loader, and it is likely that similarities will be found in earlier stages in the

replication process, when the DNA-unwinding helicase proteins are loaded.

Bowman *et al.*² and Miyata *et al.*³ now offer new insight into the clamp-loading reaction, through structural analyses of the complexes formed between a sliding clamp and its cognate clamp loader in yeast² and in the archaeon *Pyrococcus furiosus*³. The clamp loader in question is named replication factor-C (RFC). In yeast, RFC has one large subunit and four smaller components with related sequences; in archaea (microbes that have a simpler replication machinery than yeast), it has one large and four identical small subunits. The sliding clamp, meanwhile, is named proliferating cell nuclear antigen (PCNA) and consists of three identical subunits.

It has been known for many years that RFC can, in the presence of the high-energy cofactor adenosine triphosphate (ATP), mediate the loading of PCNA onto DNA^{5,6}. The RFC clamp loader then hydrolyses the ATP, resulting in the clamp being handed off to DNA polymerase. Taking advantage of the fact that a stable complex can be formed between DNA, PCNA and RFC in the presence of a poorly hydrolysable ATP analogue, Miyata *et al.*³ were able to purify such a complex from *Pyrococcus* and subject it to electron microscopy followed by three-dimensional reconstruction.

The resulting image is intriguing. The structure resembles a horseshoe stacked on a doughnut (Fig. 1), with the doughnut corresponding to the clamp, PCNA, and the horseshoe to the clamp loader, RFC. It seems as though the large subunit of RFC contacts one subunit of PCNA, and two or three of the RFC small subunits contact the remaining two PCNA subunits. DNA passes through the open channel in the horseshoe of RFC. Thus, this structure presumably corresponds to the loaded complex, just before PCNA binds DNA polymerase.

Meanwhile, Bowman *et al.*² describe a stunning, high-resolution atomic structure of yeast RFC in complex with PCNA, again in the presence of a poorly hydrolysable ATP analogue. In this case, DNA was not present in the crystallized complex itself but was incorporated in the three-dimensional reconstruction. As in the archaeal image, RFC sits on top of PCNA, with the large subunit contacting one PCNA subunit. But there are some important differences in the interactions of the RFC small subunits. Most significantly, instead of lying flat on PCNA, contacting the remaining two PCNA subunits, the RFC small subunits spiral up and away from the plane of the PCNA ring and contact only one further PCNA subunit (Fig. 1). Remarkably, the pitch of the spiral produced by RFC matches the geometry of the DNA double helix, and the authors present a persuasive model for how the subunits

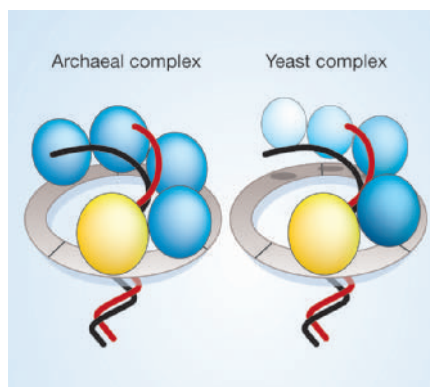


Figure 1 Clamping down on DNA. The new papers^{2,3} describe the structures of the yeast² and archaeal³ clamp-clamp-loader complexes; the similarities and differences are depicted here. The sliding clamp, PCNA, is in grey. The clamp loader, RFC, is shown as a series of coloured ovals, with the large subunit in yellow and the small subunits in blue. The yeast small subunits have distinct but related sequences (indicated by graded colours); the archaeal small subunits are identical. The predicted path of DNA through the complex is also shown, with the template strand in black and the new strand in red.

might contact the DNA so as to direct loading.

So why is there this difference? Setting aside trivial explanations of species differences or methods of sample preparation, could it be that these two complexes represent distinct stages in the clamp-loading process? Might the archaeal structure represent the initial complex and the yeast one a later stage? If so, might the differences arise

by virtue of an isomerization of RFC that occurs before ATP is hydrolysed and PCNA is handed over to DNA polymerase? And could the fact that one subunit of PCNA is exposed in the yeast structure therefore be indicative of the route that the DNA polymerase takes to access the ring? Finally, it remains to be seen how RFC actually opens the PCNA ring. Might the spiral of RFC subunits indicate that this clamp loader pries PCNA open by pulling a subunit or subunits out of the plane of the ring, popping it open like a lock washer? Or does RFC wrench PCNA open within the plane of the ring, creating a horseshoe structure that mirrors its own?

These structures provide exciting snapshots of the manner in which RFC and PCNA interact. Although small differences have been identified, conserved themes in this essential reaction are coming to the fore. With continuing efforts to study other stages of clamp loading, a full molecular view of the process seems to be on the horizon.

Michael A. Trakselis and Stephen D. Bell are in the Medical Research Council Cancer Cell Unit, Hutchison MRC Research Centre, Hills Road, Cambridge CB2 2XZ, UK.
e-mails: sb419@cam.ac.uk
mt339@cam.ac.uk

1. Watson, J. D. H. & Crick, F. H. C. *Nature* **171**, 737–738 (1953).
2. Bowman, G. D., O'Donnell, M. & Kuriyan, J. *Nature* **429**, 724–730 (2004).
3. Miyata, T. *et al.* *Nature Struct. Mol. Biol.* doi:10.1038/nsmb788 (2004).
4. Warbrick, E. *BioEssays* **22**, 997–1006 (2000).
5. Tsurimoto, T. & Stillman, B. *Proc. Natl Acad. Sci. USA* **87**, 1023–1027 (1990).
6. Gomes, X. V. & Burgers, P. M. J. *Biol. Chem.* **276**, 34768–34775 (2001).

Laser physics

Fantastic plastic

Ifor D. W. Samuel

Plastics are ubiquitous, thanks to the cheapness and versatility of these materials. Now plastic lasers are in prospect, battery-operated for low-cost communication and display applications.

Nearly all plastics are electrical insulators, but one class of plastics is different. These are conjugated polymers, whose discovery was celebrated with the award of the 2000 Nobel Prize in Chemistry. They differ from other polymers in having a backbone of alternating single and double bonds, and this difference in structure makes them semiconductors. Conjugated polymers are therefore interesting materials, with their unique combination of semiconducting properties and scope for simple fabrication and shaping. Lasers are a promising application, and now Reufer *et al.*¹, writing in *Applied Physics Letters*, report fresh progress towards the creation of low-cost polymer lasers.

A major breakthrough in the development of semiconducting polymers was the

discovery that when a voltage is applied to a thin layer of one of these materials, light can be emitted². Light-emitting diodes using this effect are now the basis of a modern flat-panel display technology. Other polymer optoelectronic devices have followed, including polymer solar cells, optical amplifiers and lasers, although further development is necessary before they reach the market. Polymer lasers are attractive as light sources for several reasons. Polymers can emit light across the visible spectrum, and so wavelengths can be generated that are not readily available from other lasers. The polymers have broad spectra, meaning that a polymer laser can be tuned over a range of wavelengths. Polymer lasers should also be simple to make, can be flexible, and should

NATURE

100 YEARS AGO

The most important archaeological event reported from Egypt during the last excavation season (1903–4) is the discovery by Prof. Naville, of the University of Geneva, and Mr. H. R. Hall, of the British Museum, of the most ancient temple at Thebes... It is the funerary temple or mortuary chapel of the most distinguished monarch of the eleventh dynasty, Nebkherurā Mentuhotep, who reigned about 2,500 B.C.... So far as platform, ramp, and colonnades are concerned, this is precisely the arrangement of the great temple of Queen Hatshepsut, or Hatasu, to the north... The curious plan of the great temple has puzzled archaeologists and architects from Wilkinson's time to the present day. Whence this curious arrangement of platforms, inclined planes, and colonnades, so totally unlike anything else in Egypt? Various theories have been propounded, but it is only now that the solution has been found, owing to the discovery of the temple of Nebkherurā. Colonnades, platforms and ramps are then a feature of the older temple-architecture of Egypt; they were, at the time of the eighteenth dynasty, when the great temple of Hatshepsut was built, old-fashioned, archaic, but it is evident that the great temple is, as far as its main arrangements are concerned, a mere enlarged copy of the thousand-year older temple at its side.

From *Nature* 16 June 1904.

50 YEARS AGO

The existence of a large number of discrete sources of extraterrestrial radio radiation (radio stars) has now been established. For the majority of the sources there is no information about their distance... The present communication describes a new method for the measurement of the distance of radio stars in the Galaxy, and gives preliminary results for the two intense sources in Cygnus and Cassiopeia. The method depends upon the absorption of the radiation from the radio stars by interstellar hydrogen at a frequency of about 1,420 Mc./s., corresponding to the spectral line of the ground-state of the neutral hydrogen atom. If the radiation from the source traverses neutral hydrogen in the Galaxy, it will suffer absorption at the line frequency, whereas its emission at neighbouring frequencies will be unaffected... Thus from observations of the spectral distribution and the magnitude of the absorption, the distance of the radio star can be estimated. D. R. W. William & R. D. Davies
From *Nature* 19 June 1954.

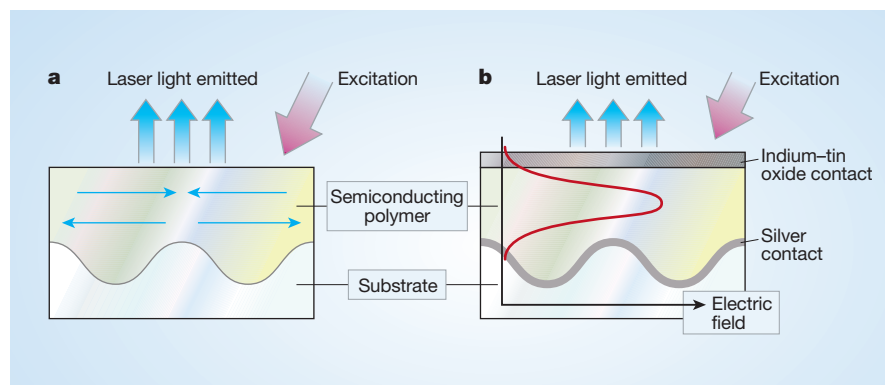


Figure 1 Low-cost lasers from plastic. a, A laser consists of a light-amplifying material and a resonator. Polymer (plastic) semiconductors can be used to amplify light, with a corrugated structure acting as a resonator. The corrugation diffracts light travelling to the left back towards the right, and vice versa. Hence, light is sent backwards and forwards through the amplifying medium. Typically, the excitation is provided by another laser. It would be more convenient, however, to make polymer lasers that operate using electrical excitation — say, from a battery. b, Reufer *et al.*¹ have shown how to apply electrical contacts to a semiconducting polymer laser without spoiling its optical properties. With a thick semiconducting polymer layer in the laser, the electric field of light in the polymer is weak at the contacts. So absorption (hence, loss) of light at the contacts is low and the laser works effectively (although still, for now, with laser excitation).

be suitable for a wide range of applications, from plastic optical circuits to biological screening and assays.

Lasers consist of two key components: the gain medium and a resonator. The gain medium is a material in which light is amplified. In most lasers, this would be a crystalline inorganic material, such as ruby or gallium arsenide (the latter is commonly used in CD players and laser pointers). Light passes backwards and forwards through the gain medium, thanks to the feedback generated by the resonator. In plastic lasers, a semiconducting polymer is used as the gain medium^{3–5}. Light can be passed backwards and forwards through the polymer with mirrors^{3,4}, but in most polymer-laser work, including that of Reufer *et al.*¹, a corrugated gain medium is used instead (Fig. 1a). The corrugation acts as a diffraction grating that diffracts light travelling in one direction back in the opposite direction, thereby creating feedback and enabling lasing to occur. As there are no mirrors in these lasers, they are compact, robust and exceptionally easy to align.

A laser needs energy in order to operate. This can be supplied in two ways: either optically (from flash-lamps or another laser) or electrically. There is a minimum energy, the threshold, required for operation. Above threshold, the gain (or amplification) exceeds all the losses in the device, and lasing begins. At present, all polymer lasers are optically excited, or 'pumped'. However, as mentioned above, light emission from semiconducting polymers can be induced electrically, and an electrically pumped laser would be much more convenient.

There are three main reasons why it has so far been possible to make only optically pumped polymer lasers. The first is that, in

order to reach the threshold typical of existing lasers, a higher current density would be required in a polymer laser than most semiconducting polymers can readily withstand. The second is that an electrically pumped polymer laser would require electrical contacts. These would normally be metal and would introduce substantial additional losses, raising the threshold even further. Finally, the electrical charges injected would also absorb light, leading to increased losses and higher threshold.

The first problem can be solved by using pulsed excitation. Reufer *et al.*¹ have addressed the second problem — the losses associated with electrical contacts. A few groups have suggested possible solutions^{5–8}, and lasing has been achieved in the presence of a metal contact but with a considerably increased threshold⁹. Now Reufer *et al.* have demonstrated a way in which a metal contact can be applied without a threshold increase (Fig. 1b). The losses depend on the electric field distribution through the polymer layer. If the polymer layer is made thicker, this reduces the strength of the electric field at the metal contact, thereby reducing the associated losses. The metal contact used was silver, but the approach should be more generally applicable. To allow the light out of the laser, Reufer *et al.* used a very thin (20 nm) layer of indium–tin oxide, a transparent conductor, as the other electrical contact. This additional layer only slightly increased the threshold.

It should be noted that the polymer laser was optically pumped (that is, energy was supplied by another laser). But this latest work does show a way of dealing with the losses associated with electrical contacts, and hence is a significant step towards an electrically pumped polymer laser. There

are limitations to Reufer and colleagues' laser, however. In particular, the amount of current that could be passed through the device is limited for several reasons: silver contacts are not ideal for injecting charge effectively into semiconducting polymers; the thicker polymer film used in this set-up makes charge injection and transport more difficult; and the electrical conductivity of the thin indium–tin oxide film is relatively poor, limiting the area of such lasers and increasing their operating voltage. Overall, however, Reufer and colleagues' results¹ bring low-cost, battery-operated, visible lasers a step closer, and will stimulate renewed interest in plastic lasers. ■

Ifor D. W. Samuel is at the Organic Semiconductor Centre, School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK.
e-mail: idws@st-andrews.ac.uk

1. Reufer, M. *et al.* *Appl. Phys. Lett.* **84**, 3262–3264 (2004).
2. Burroughes, J. H. *et al.* *Nature* **347**, 539–541 (1990).
3. Moses, D. *Appl. Phys. Lett.* **60**, 3215–3216 (1992).
4. Tessler, N. *Adv. Mater.* **11**, 363–370 (1999).
5. McGehee, M. D. & Heeger, A. J. *Adv. Mater.* **12**, 1655–1668 (2000).
6. Frolov, S. V. *et al.* *Appl. Phys. Lett.* **72**, 2811–2813 (1998).
7. Kozlov, V. G. *et al.* *IEEE J. Quantum Electron.* **36**, 18–26 (2000).
8. Stehr, J. *et al.* *Adv. Mater.* **15**, 1726–1729 (2003).
9. Andrew, P., Turnbull, G. A., Samuel, I. D. W. & Barnes, W. L. *Appl. Phys. Lett.* **81**, 954–956 (2002).

Neurobiology

Why voles stick together

Evan Balaban

The tendency for animals to form social bonds after sexual activity varies greatly from species to species. Work with voles illuminates a molecular pathway in the brain that influences such differences.

A report on page 754 of this issue continues a fascinating line of inquiry into the basic brain mechanisms that contribute to social behaviour. There, Lim and colleagues¹ show that increasing the expression of a single protein in a particular brain region of male meadow voles makes them more socially cohesive — rather like their close relation the prairie vole.

There is much that researchers would like to know about the social behaviour of animals (including ourselves, as discussed on page 705 of this issue²). What, for instance, are the brain mechanisms that mediate the formation of bonds between individuals? How do these mechanisms change within and between populations and species over evolutionary time? The humble vole has provided a wonderful opportunity to pursue these issues.

There are many species of vole, and they exhibit markedly different patterns of social attachment. For instance, male prairie voles in captivity are likely to form preferential associations with one female, defined by close physical contact, choosing to spend more time with her when given equal access to several females, and keeping other males away from her³. These preferential associations form most readily after sexual activity. Captive male prairie voles also tend to interact with and care for their young. (Both traits vary in degree among wild populations⁴.) Captive males of the closely related meadow vole, by contrast, exhibit weaker 'pair-bonding' with females (again, this is variable in the wild⁵), and give little attention to young.

Studies of the social organization, behavioural ecology and hormones of voles^{6,7} have

linked such differences in 'affiliative behaviour' to variation in the expression patterns of the receptors for two related signalling molecules, oxytocin and arginine vasopressin (vasopressin for short)³. These molecules are released in the brain after sex, and are also involved in other reproduction-related behaviours and in unrelated functions such as water retention and stress.

Notably, the vasopressin 1a receptor (V1aR) is expressed in greater numbers in the ventral pallidum region of the forebrain in male prairie voles than in male meadow voles⁸. Although the exact relationships of the ventral pallidum to other brain regions and to behaviour are still unclear, the neighbouring nucleus accumbens is part of the brain's 'reward system', which signals that a particular behaviour is worth doing again. In prairie voles, manipulation either of signals mediated by the neurotransmitter dopamine in the nucleus accumbens⁹, or of vasopressin signalling within (at least in part) the ventral pallidum^{10,11}, can substitute for or block the effects of sexual activity on preferential association — showing the importance of both regions in forming attachments. The expression of dopamine receptors shows little natural variation between species, however, hinting that the key to the differing behaviour of prairie and meadow voles might lie in the variation in vasopressin receptors.

Lim *et al.*¹ now take us beyond correlations in the study of these receptors, by experimentally manipulating the expression of V1aR. The authors wanted to find out whether adult male meadow voles that have prairie-vole-like expression of

vasopressin receptors also show prairie-vole-like social behaviour. They discovered that, as a group, male meadow voles that have more V1aRs in the ventral pallidum spend more time both with their mates and near juveniles than do controls. They do not, however, engage in paternal care (consistent with previous suggestions that these traits can be dissociated in prairie voles³). Lim *et al.* also show that, as in prairie voles, preferential association in the experimentally manipulated meadow voles is prevented by prior blockage of dopamine receptors. A similar study of adult male prairie voles with above-average V1aR levels¹² found a potentiation of preferential association with females, and increased time spent in proximity to unfamiliar juveniles.

So Lim *et al.* argue that, in male prairie voles, the sexually induced release of vasopressin triggers its receptors in the ventral pallidum, which somehow enable the intrinsically rewarding sensations experienced after copulation to become reliably associated with the individual features of a particular female, such as her odour. (It is unclear whether vasopressin signalling simply causes familiar animals to spend more time close to one another, allowing the association with the rewarding effects of sex to occur more reliably, or whether it directly causes changes in the inputs to, or function of, the brain reward systems.) Male meadow voles, by contrast, form weaker social bonds because of a lower level of signalling in the relevant brain region. But because the subsequent molecular and neuronal circuitry is highly similar in these species, changing the expression of this one receptor can profoundly alter the circuit's function — implying that evolutionary selection can act on this single molecule to produce major changes in social behaviour.

If the V1aRs are indeed the adjustable nozzle atop a social-glue dispenser in the mammalian brain, these results could have wider significance for understanding social behaviour — and some of its dysfunctions, as seen, for instance, in autism — in humans and animals. Caution is warranted on three fronts, however. First, several studies^{4,13,14} indicate that changes in V1aR expression alone might not fully account for naturally observed differences in pair-bond formation in voles. Individuals from populations and species with similar V1aR levels but different behaviour will be valuable in the search for other molecular and cellular components of the social glue.

Second, vasopressin signalling mediates many other reproductively related and unrelated functions. So the experimental change in the meadow-vole ventral pallidum might detrimentally affect other brain functions in manipulated individuals — unless they have additional, compensatory genetic modifications, as prairie voles evidently do.

An understanding of the wider evolutionary significance of this manipulation must await a fuller documentation of its physiological consequences.

Finally, it has been suggested^{11,3,15} that variation in the expression of the V1aR gene, and hence evolutionary changes in vole social behaviour, can be attributed solely to variation in the gene's regulatory DNA sequences. But perhaps we should not be so quick to leap to this conclusion. For instance, mice engineered to include a prairie-vole V1aR gene (plus control sequences) fail to display prairie-vole-like levels of V1aR expression in the ventral pallidum¹⁶. Similarly, control sequences from the prairie-vole V1aR gene do not yield higher expression levels in rat brain cell lines than do such sequences from montane voles (which are similar to meadow voles)¹⁵. Gene variation in signalling or regulatory molecules that interact with the V1aR control region are also likely to be important.

Understanding the role of genetic variation in the evolution of any trait requires knowledge of the major genes involved, their distribution among individuals, the distribution of genetically defined individuals in different environments, and the dependence of the trait on the environment for each individual and at each stage of development. This is a tall order, especially for behavioural characteristics. But the research on voles

gives us hope that classical comparative studies of natural populations, judiciously coupled with modern molecular and cellular neurobiology, will continue to provide insights into the relationships between genes, brain-cell collectives, ecology and chance in social behaviour. ■

Evan Balaban is in the Behavioural Neurosciences Program, McGill University, 1205 Avenue Docteur Penfield, Montreal H3A 1B1, Canada.
e-mail: evan@psych.mcgill.ca

1. Lim, M. M. *et al.* *Nature* **429**, 754–757 (2004).
2. Konner, M. *Nature* **429**, 705 (2004).
3. Insel, T. R. & Young, L. J. *Nature Rev. Neurosci.* **2**, 129–135 (2001).
4. Cushing, B. S., Martin, J. O., Young, L. J. & Carter, C. S. *Horm. Behav.* **39**, 48–58 (2001).
5. Parker, K. J., Phillips, K. M. & Lee, T. M. *Anim. Behav.* **61**, 1217–1226 (2001).
6. Dewsbury, D. A. *Nebraska Symp. Motiv.* **35**, 1–50 (1988).
7. Carter, C. S., DeVries, A. C. & Getz, L. L. *Neurosci. Biobehav. Rev.* **19**, 303–314 (1995).
8. Young, L. J., Winslow, J. T., Nilsen, R. & Insel, T. R. *Behav. Neurosci.* **111**, 599–605 (1997).
9. Aragona, B. J., Liu, Y., Curtis, T., Stephan, F. K. & Wang, Z. *J. Neurosci.* **23**, 3483–3490 (2003).
10. Lim, M. M. & Young, L. J. *Neuroscience* **125**, 35–45 (2004).
11. Lim, M. M., Murphy, A. Z. & Young, L. J. *J. Comp. Neurol.* **468**, 555–570 (2004).
12. Pitkow, L. J. *et al.* *J. Neurosci.* **21**, 7392–7396 (2001).
13. Cushing, B. S., Okorie, U. & Young, L. J. *J. Neuroendocrinol.* **15**, 1021–1026 (2003).
14. Phelps, S. M. & Young, L. J. *J. Comp. Neurol.* **466**, 564–576 (2003).
15. Hammock, E. A. D. & Young, L. J. *Mol. Biol. Evol.* **21**, 1057–1063 (2004).
16. Young, L. J., Nilson, R., Waymire, K. G., MacGregor, G. R. & Insel, T. R. *Nature* **400**, 766–768 (1999).

science. For example, if entangled particles were distributed throughout various sectors of a quantum computer, then quantum teleportation could provide a means for distant quantum bits (or qubits) to interact without the requirement of physical proximity — effectively ‘quantum wiring’, with desirable scaling properties. In addition, disposable quantum software could be delivered from a remote location using a generalized form of quantum teleportation to enhance the capabilities of rudimentary quantum hardware⁷.

Remarkably, the two groups^{5,6} have used quite different techniques for achieving teleportation, and yet both reach very similar values of so-called fidelity. Fidelity is a figure of merit that quantifies how well the quantum state that appears in the second ion after teleportation resembles the original quantum state; fidelity is 1 in the ideal case. Both teams report values around 0.75, which exceeds the ‘classical’ value of 2/3 that can be reached without quantum entanglement. For classical teleportation, the original quantum state is simply measured, and a new quantum state recreated by using only the classical information obtained from the measurement.

Both experiments have thereby reached the milestone of unconditional, or deterministic, teleportation of atomic qubits. The initial quantum state is prepared on demand, then teleported from one ion to another with high efficiency at the push of a button (which in fact triggers a computer-controlled array of complex operations). The teleported state is then available for further experiments. Such bona fide teleportation of quantum bits, following the original proposal of Bennett *et al.*¹, has not been achieved before — not in experiments with polarization states of light, and certainly not for any material system. The only other setting in which deterministic teleportation has been realized is that of continuous quantum variables (roughly, the amplitude and phase of a beam of light)⁴.

In terms of the actual physical systems, Riebe *et al.*⁵ employ ground and metastable states of trapped calcium ions as qubits; Barrett *et al.*⁶ utilize two ground states in the hyperfine structure of beryllium ions. As for the implementation of quantum operations, the two experiments differ in several important aspects. First, crucial elements of both teleportation and quantum computing are joint operations for two qubits that cannot be performed by simply manipulating the qubits separately. Such two-body interactions are required for the creation of entanglement between two ions (step 1 in Fig. 1), and for the implementation of joint or Bell-state measurements (step 3). Riebe *et al.* use a version of the Cirac–Zoller two-qubit gate⁸, which relies on the common centre-of-mass motion of the ions. Barrett

Quantum physics

Push-button teleportation

H. J. Kimble and S. J. van Enk

Two groups have succeeded in teleporting quantum states between different atoms — a spectacular advance in the quest to achieve quantum computation.

In 1993, Charles Bennett and colleagues described a remarkable protocol for transporting a quantum state from one location to another¹, a protocol that succeeds even when the quantum state is completely unknown at the respective sites. Such quantum teleportation makes use of an extraordinary quantum resource, namely entanglement between two systems. Moreover, it also requires ordinary classical information obtained by performing a joint measurement on the system that carries the quantum state to be teleported and one component of the entangled state (as outlined in Fig. 1). Strangely, neither classical nor quantum channels individually carry any information about the quantum state, leading to the characterization of teleportation as the disembodied transport of quantum states. Initial experimental demonstrations of quantum teleportation, from 1997 onwards, involved the quantum states of beams of light^{2–4}. Now, in a

landmark advance, two teams have achieved teleportation for the quantum states of massive particles^{5,6}.

As described on pages 734 and 737 of this issue, Riebe *et al.*⁵ and Barrett *et al.*⁶ have generated coherent superpositions of two internal states for a single trapped ion (P in Fig. 1), and have teleported these quantum states to a second ion (B), with the help of a third, auxiliary ion (A). The import of these experiments goes well beyond the demonstration of teleportation *per se*, because both schemes incorporate many complex procedures that are required for scalable quantum computing. Indeed, the ion-trap set-up is generally considered one of the most promising implementations for quantum computing, as is once again confirmed by these experiments.

Moreover, quantum teleportation has emerged as an essential operation for diverse tasks in quantum information

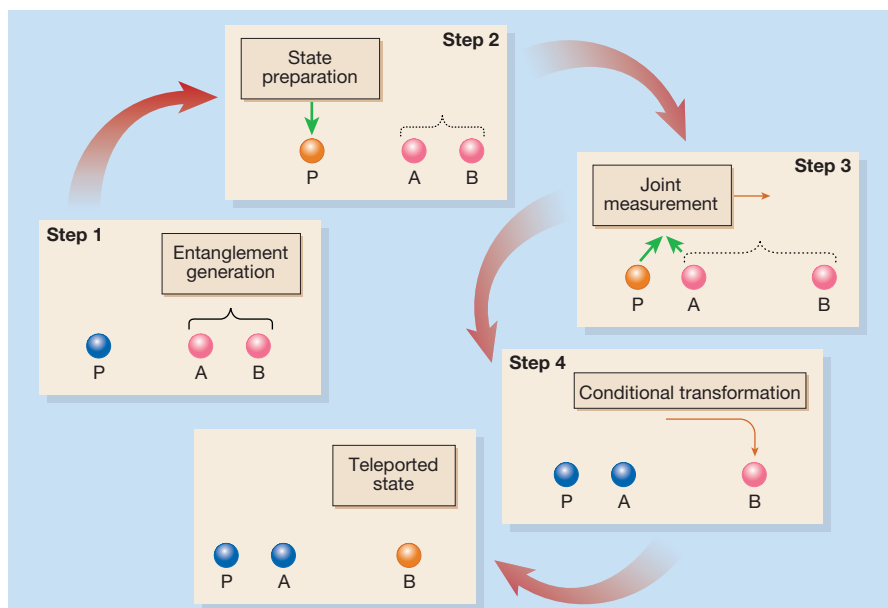


Figure 1 Quantum teleportation, step by step. Although the details of their experiments differ, Riebe *et al.*⁵ and Barrett *et al.*⁶ have followed the protocol suggested by Bennett *et al.*¹ to achieve deterministic teleportation of a quantum state between trapped ions. First, an entangled state of ions A and B is generated, then the state to be teleported — a coherent superposition of internal states — is created in a third ion, P. The third step is a joint measurement of P and A, with the result sent to the location of ion B, where it is used to transform the state of ion B (step 4). The state created for P has then been teleported to B.

et al. adopt a recently developed geometric method to perform two-qubit gating⁹.

A second difference concerns how the authors addressed individual ions for manipulating quantum states, including projective measurements. Riebe *et al.* are able

to address any specified target ion using tightly focused laser beams and have developed a technique to 'hide' the remaining ions from the target ion's fluorescence by changing their internal states so that they are insensitive to the fluorescent light. Barrett *et al.*

have developed the capability to move groups of ions selectively to separate zones in a segmented trap, thereby isolating any target ion while still maintaining entanglement within the system.

The details of implementation aside, these two experiments represent a magnificent confluence of experimental advances, ranging from precision spectroscopy and laser cooling to new capabilities for controlled two-body interactions. The techniques developed and employed by these groups will no doubt prove important in the quest to build large-scale quantum computers based on trapped ions. Indeed, the fact that such diverse procedures performed so superbly in two separate laboratories attests to the flexibility and great potential of ion trapping for processing quantum information.

H. J. Kimble is in the Norman Bridge Laboratory of Physics, California Institute of Technology, Pasadena, California 91125, USA.

e-mail: hjkimble@its.caltech.edu

S. J. van Enk is at Bell Labs, Lucent Technologies, 600–700 Mountain Avenue, Murray Hill, New Jersey 07974, USA.

1. Bennett, C. H. *et al.* *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
2. Bouwmeester, D. *et al.* *Nature* **390**, 575–579 (1997).
3. Boschi, D., Branca, S., De Martini, F., Hardy, L. & Popescu, S. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
4. Furusawa, A. *et al.* *Science* **282**, 706–709 (1998).
5. Riebe, M. *et al.* *Nature* **429**, 734–737 (2004).
6. Barrett, M. D. *et al.* *Nature* **429**, 737–739 (2004).
7. Gottesman, D. & Chuang, I. L. *Nature* **402**, 390–393 (1999).
8. Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
9. Leibfried, D. *et al.* *Nature* **422**, 412–415 (2003).

Atmospheric chemistry

Fire and ice

Small air bubbles, trapped in glaciers and ice caps, have been a priceless source of information about the history of Earth's atmosphere: the record now goes back some 800,000 years. But, for a gas to be preserved in this frozen archive of atmospheric history, it must be chemically inert, and so there can be no direct record of the short-lived species, such as ozone (O_3) and the hydroxy radical (OH), that drive the key chemical processes in the atmosphere.

Writing in the *Journal of Geophysical Research* (doi:10.1029/2003JD004218; 2004), Becky Alexander *et al.* suggest that a little-known peculiarity of isotope chemistry, known as 'mass-independent fractionation', may offer a glimpse into the history of vegetation burning and its effects on oxidation reactions in the atmosphere.

The fractionation of isotopes that occurs during most chemical and

physical processes is usually proportional to the mass difference between the isotopes. But a few reactions, notably the formation of ozone, also lead to isotope fractionations that are not proportional to the isotope mass. The resulting anomaly in the isotopic composition of atmospheric ozone is passed on to the sulphate and nitrate formed from aqueous-phase reactions between O_3 and SO_2 and NO_x respectively. Other processes that lead to sulphate and nitrate, such as gas-phase reactions involving the hydroxy radical, do not show this anomaly.

As a result, we can tell the relative contributions of liquid-phase oxidation by O_3 and gas-phase oxidation by OH from the isotopic composition of sulphate and nitrate trapped in the layers of ancient ice in Greenland and Antarctica. Alexander *et al.* have identified a large isotope anomaly in Greenland ice that dates from the

period 1830–1900, and it turns out to correlate perfectly with indicators of biomass burning — formate and ammonium — in the same ice layers. They propose that this anomaly is the consequence of elevated levels of ozone that stemmed from air pollution from the fires that the pioneers used to clear North America for agriculture in the nineteenth century.

These data cannot yield quantitative conclusions about the amounts of oxidants in the pre-industrial atmosphere. They can, however, serve as constraints for models that integrate information about the emissions that are associated with different types of land use and their reactions in the atmosphere. Such models are used for analysing and predicting the human impact on atmospheric composition. In the tropics, high concentrations of water vapour and solar ultraviolet radiation combine to produce the



highest OH concentrations worldwide, making this area the most important region for atmospheric oxidation. So, for further understanding of those processes, we might want to look with added urgency for isotopic clues in the few and rapidly vanishing ice caps on mountains in the tropics (pictured, Mount Huascarán in Peru, in 1978).

Meinrat O. Andreae

Meinrat O. Andreae is at the Max Planck Institute for Chemistry, PO Box 3060, D-55020 Mainz, Germany.

e-mail: andreae@mpch-mainz.mpg.de

Obituary

Carl-Ivar Brändén (1934–2004)

When Max Perutz and John Kendrew solved the structures of the oxygen-delivering proteins haemoglobin and myoglobin in the 1950s, the time was clearly ripe for more extensive studies of the molecular structures of biological systems. Thus it was that, in 1962, the UK Medical Research Council opened the splendid Laboratory of Molecular Biology (LMB) just south of Cambridge. Carl-Ivar Brändén, a mathematics and chemistry student from Uppsala, Sweden, was one of the first postdoctoral students at the new laboratory, from 1962 to 1963. His PhD work in Uppsala had dealt with the crystallography of inorganic metallic salts, but his experience at the LMB became decisive in shaping his future career, in which he used structural studies to elucidate protein function. Brändén died on 28 April 2004, just two weeks short of his seventieth birthday.

Calle, as he preferred to be called, was born in 1934 in a tiny village in Lapland. He spent his first six years at school under the supervision of his father, who was the local teacher. The village was poor; with nine months of winter, the climate was even worse. His determination to get away made him study hard, and he was subsequently awarded a state fellowship for university studies in Uppsala, where he also gained his PhD. He retained a lifelong interest in nature, however, and loved to pick mushrooms and cloudberries.

In his university years, Brändén began by studying mathematics and physics, but quickly dropped the physics, opting for chemistry instead. With this background, he was ideally placed to contribute to the computational refinement of the X-ray structures of inorganic molecules during his PhD with chemist Ingvar Lindqvist. Crystallographic refinement is used to make a rough atomic model fit the data better. At the time, there was a dearth of computer programs capable of dealing with X-ray diffraction data — a fact that prompted Brändén to write his first program for the early Swedish electronic computer, BESK, using the method of least squares. The program was used for more than ten years by the entire Scandinavian crystallographic community.

A course in biochemistry stirred Brändén to switch allegiances once again, as he became keen to apply his knowledge of computing and crystallography to biological molecules. He obtained a year's placement at the LMB, where, together with one of us (K.C.H.), he engaged in writing a



From mathematics and chemistry to computing and structural biology

computer program for refining the structure of myoglobin. This protein has 1,500 atoms, and no one had ever attempted to refine a system that big before. The method involved treating each peptide group as a planar unit, to reduce the number of parameters. In those days, one cycle of refinement required 16 hours of computing time, at a cost of £3,000 (US\$5,500). Although the method worked, the cost was prohibitive, and the structure of myoglobin was not refined for another 20 years.

When he returned to Sweden in 1963, Brändén was appointed an associate professor at the Swedish University of Agricultural Sciences in Uppsala. There, his first problem was to solve the structure of the alcohol dehydrogenase enzyme, which is involved in decomposing alcohol. Funding was scarce, and the first crystals were made in the kitchen refrigerator. Eventually, however, Brändén and his colleagues were able to postulate a detailed mechanism of action, based on the structures of several complexes in different conformations. Through pioneering studies of allosteric effects and the binding of coenzymes and substrates to enzymes, his interest in integrating structure with function and molecular biology continued into the 1980s.

When the Swedish Research Council awarded him a full professorship in molecular biology at the University of Agricultural Sciences in 1970, Brändén was ready to take on a new challenge. He decided to tackle the structure of Rubisco, the enzyme that catalyses the initial fixation of carbon dioxide during photosynthesis in green plants. Despite some initial opposition — Rubisco, being a large, multi-subunit protein complex, was thought to be too difficult a subject for crystallography — Brändén and his group

finally published the structure in 1986. The structure represented a turning point in the study of this enzyme. Not only did it provide the first detailed picture of Rubisco's active site, it also revealed the so-called fold of the large subunit of the Rubisco molecules found in all plants, aiding their interpretation.

Around this time, Brändén realized that protein structures would become an ever more important tool in the molecular life sciences. So in 1987 he and John Tooze, who ran the European Molecular Biology Organization, began writing a textbook to introduce biologists to the concepts of structure determination. The result was *Introduction to Protein Structure*, published in 1991. With its superb pictures and accessible text, the book quickly acquired a strong following among graduate students and newcomers to the field.

Brändén also had a marked impact on science policy in Sweden. He became one of the scientific advisers to the Swedish government, for instance, and served nine years on the Nobel Committee for Chemistry. Many international appointments followed, culminating with the directorship of research in life sciences at the European Synchrotron Radiation Facility in Grenoble, France, from 1992 to 1997. While there, Brändén was responsible for building up experimental facilities for chemistry, biology and medicine. He realized that this facility provided a prime opportunity for extending the frontiers of structural biology, by facilitating the determination of the structures of large molecular complexes, and enabling the dynamics of biochemical reactions to be studied in the microsecond time range by time-resolved crystallography.

Throughout his life, Calle continued to share his views on molecular biology and structure, as well as on producing a creative research environment. With his open-door policy, he also inspired and supported many young scientists and students. Calle was not only generous but also knowledgeable, determined and full of humour — characteristics that can perhaps be put down to his tough childhood years. Despite spending his final 18 months under treatment for lung cancer, from which he died, he remained optimistic and quite satisfied with the richness of his personal and professional life.

Ingemar Ernberg & Kenneth C. Holmes

Ingemar Ernberg is at the Karolinska Institutet, Stockholm S-171 77, Sweden.

e-mail: ingemar.ernberg@mtc.ki.se

Kenneth C. Holmes is at the Max Planck Institute for Medical Research, Heidelberg 69120, Germany.

e-mail: holmes@mpimf-heidelberg.mpg.de

Epilepsy

Mice with a cleaning issue

Genes Dev. **18**, 1397–1412 (2004)

In their studies of body-clock mechanisms in mice, Frédéric Gachon and colleagues have identified a set of genes that, when deleted, give rise to epilepsy. The researchers studied mutant mice that lack some or all of a set of three genes that encode a family of proteins called PAR bZip (proline- and acidic-amino-acid-rich basic leucine zipper). Concentrations of these proteins vary by up to 50-fold in the liver and parts of the brain during the 24-hour cycle.

In their attempts to decipher the functions of this family, the researchers noticed that death rates among mice lacking all three genes (triple-knockout mice) were markedly higher than those among the other knockout mice. Intriguingly, death rates were highest on Mondays and Thursdays, when the lab was cleaned. Under controlled conditions, Gachon *et al.* discovered that the noise of a vacuum cleaner induced spasms in the triple-knockout mice, showing that they had epilepsy.

The authors propose a mechanism that may account for epilepsy in mice lacking PAR bZip family members. These proteins regulate a gene involved in activating vitamin B6. The activated vitamin in turn plays a crucial role in neurotransmitter metabolism, and the brains of triple-knockout mice indeed showed altered levels of various neurotransmitters. **Michael Hopkin**

Traffic flow

Cruising through congestion

Phys. Rev. E **69**, 066110 (2004)

Vehicles fitted with adaptive cruise control (ACC) systems are now reaching mainstream production. These systems use radar to automatically maintain a constant, safe distance between cars, which potentially keeps traffic flowing more smoothly. L. C. Davis has investigated how a gradual deployment of ACC vehicles onto the streets will affect congestion.

Earlier research had found that with 20% of cars fitted with ACC and the remainder under manual control, high-speed traffic flowed freely without jams. Davis reports that with only 10% ACC vehicles on the road, jams still form quickly in highway traffic moving at around 90 km per hour. But at around 13% ACC coverage, changing just one manual vehicle over to ACC stops the jam forming.

Davis suggests that this discontinuity is caused by the jam's strong dependence on the precise sequence of ACC and manual vehicles in the queue. At lower speeds, there



is no such critical concentration: increasing ACC vehicles on the road gradually reduces the likelihood of jamming. However, Davis warns that because ACC vehicles maintain precise distances between each other on the road, there are no obvious safe gaps for a manual driver wanting to join the traffic flow. This may result in increased congestion around motorway slip roads, for example. **Mark Peplow**

Medicine

Tumour tracking

Proc. Natl Acad. Sci. USA
doi:10.1073/pnas.0402993101 (2004)

Around a quarter of breast tumours acquire extra copies of a gene called *HER-2*. Although these tumours are linked to poor survival, the patient's prognosis can be improved with the drug herceptin. But most doctors test for the mutation only once, in the primary breast tumour, and assume that it does not crop up later on. Songdong Meng *et al.* now challenge that assumption. They found that 9 out of 24 patients who originally lacked the *HER-2* mutation developed it later. Of four of these placed on herceptin, three showed some degree of tumour remission.

To test repeatedly for the *HER-2* mutation, the group had adapted a non-invasive method in which antibodies bind to and extract rare cancerous cells that have been shed from the tumour and are circulating in the blood. They spread these cells on a slide, and tested them first for high levels of *HER-2* protein and then for changes in the *HER-2* gene. They found that mutations in these circulating tumour cells correspond to mutations in the mother tumour. If the same holds true in larger studies, and if the technique can be automated, breast-cancer patients might be tested for this or other mutations throughout their disease, and treated appropriately. **Helen Pearson**

Cell signalling

Gut reaction

J. Biol. **3**, 11.1–11.17 (2004)

Signalling pathways that are activated by proteins of the Wnt family help to regulate cell proliferation, cell differentiation and organ formation. Tadashi Okubo and Brigid L. M. Hogan have found, for instance, that this pathway can steer cells that normally generate lung to form gut instead.

The authors studied the lungs of mouse fetuses that had been genetically altered to have artificially high levels of Wnt signalling activity. Externally the lungs looked normal, but internally they lacked the organ's typical branching, tree-like structure. They contained few mature lung cell types, but had many rapidly dividing cells that expressed a variety of genes known to regulate gut development.

The results suggest that immature lung cells can be coaxed to become intestinal cells through the effects of Wnt signalling — a finding that may help those wishing to control the behaviour of adult stem cells. It may also shed light on the medical condition Barrett's oesophagus, in which patches of cells lining the gullet become intestinal in nature. Higher levels of Wnt signalling in adult stem cells might be partly responsible for the condition, the authors suggest. **Helen R. Pilcher**

Chemistry

Under pressure

Chem. Commun. doi:10.1039/b404181j (2004)

Supercritical fluids have potential as environmentally friendly solvents in a variety of chemical reactions, because the products can be extracted from solution simply by reducing the pressure and removing the solvent as a gas. Jason R. Hyde and Martyn Poliakoff report a new method for adding hydrogen to organic molecules that starts, curiously, with reactant and supercritical solvent combined in a single compound: formic acid (HCO_2H). The pressurized liquid acid is easier to handle than high-pressure hydrogen and carbon dioxide gas.

A hot platinum catalyst breaks the molecule into a supercritical $\text{CO}_2\text{--H}_2$ mixture, which meets the organic molecules in a second catalytic chamber at a temperature of 80 °C and a pressure of 8 megapascals. The authors report good yields of various alkane products, which are collected when the pressure is released at the end of the process.

For versatility, the concentration of hydrogen in the system should be controllable. Hyde and Poliakoff show that this can be achieved by adding ethyl formate ($\text{HCO}_2\text{CH}_2\text{CH}_3$) to the original stream. After decomposition, this adds supercritical ethane (C_2H_6) and more CO_2 to the mixture, diluting the hydrogen content. **Mark Peplow**

Routing of spike series by dynamic circuits in the hippocampus

Frédéric Pouille & Massimo Scanziani

Neurobiology Section, Division of Biological Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0357, USA

Recurrent inhibitory loops are simple neuronal circuits found in the central nervous system, yet little is known about the physiological rules governing their activity. Here we use simultaneous somatic and dendritic recordings in rat hippocampal slices to show that during a series of action potentials in pyramidal cells recurrent inhibition rapidly shifts from their soma to the apical dendrites. Two distinct inhibitory circuits are sequentially recruited to produce this shift: one, time-locked with submillisecond precision to the onset of the action potential series, transiently inhibits the somatic and perisomatic regions of pyramidal cells; the other, activated in proportion to the rate of action potentials in the series, durably inhibits the distal apical dendrites. These two operating modes result from the synergy between pre- and postsynaptic properties of excitatory synapses onto recurrent inhibitory neurons with distinct projection patterns. Thus, the onset of a series of action potentials and the rate of action potentials in the series are selectively captured and transformed into different spatial patterns of recurrent inhibition.

Action potentials are the principal means of communication between neurons. It is believed that action potentials convey information through their timing and frequency^{1–3}. To process this information, the brain needs neuronal circuits able to extract and represent these temporal features out of series of action potentials. Among cortical areas, the hippocampus is one of the most accessible to neuronal circuit analysis⁴. We therefore investigated whether distinct temporal features of a series of spikes in CA1 pyramidal cells—the principal hippocampal output—are selectively captured by local neuronal circuits. Because inhibitory neurons represent the main local targets of recurrent excitatory synapses of CA1 pyramidal cell axons⁴, we monitored synaptic recurrent inhibition on CA1 pyramidal cells. Given that inhibitory synapses are distributed along the entire somato-dendritic axis of CA1 pyramidal cells⁵, we recorded from the soma and the apical dendrite simultaneously^{6–8}: $307 \pm 8 \mu\text{m}$ from the soma; range, 233–427 μm ; resting V_m soma, $-56.4 \pm 0.9 \text{ mV}$; resting V_m dendrite, $-57.9 \pm 0.9 \text{ mV}$; $n = 28$ (Fig. 1a). Recurrent inhibition was evoked by stimulating the axons of CA1 pyramidal cells at various frequencies with an electrode placed in the alveus (see Methods; Fig. 1a)^{9–12}.

Shift of recurrent inhibition along the somato-dendritic axis

At high frequencies, the onset of a series of stimuli delivered to the alveus elicited recurrent inhibition targeting the soma, whereas stimuli later in the series shifted recurrent inhibition towards more apical dendritic regions. Specifically, alveus stimuli repeated at 0.2 Hz elicited recurrent inhibitory postsynaptic potentials (IPSPs) that were $25.3 \pm 1.8\%$ smaller at the dendritic recording site than at the somatic one: $1.36 \pm 0.17 \text{ mV}$ versus $0.99 \pm 0.12 \text{ mV}$; $P = 2 \times 10^{-6}$; $n = 24$ (Fig. 1b, c). In contrast, with series of two to four stimuli at 100 Hz (delivered every five seconds), inhibition became progressively larger in the dendrites than in the soma: $16.4 \pm 5.3\%$ larger after four stimuli; $1.52 \pm 0.21 \text{ mV}$ in the soma versus $1.73 \pm 0.21 \text{ mV}$ in the dendrite; $P = 0.011$; $n = 12$ (Fig. 1b, c). This shift from the soma to the dendrite is clearly illustrated by the concomitant decrease and increase of the slopes of IPSPs recorded in the soma and in the dendrites, respectively (Fig. 1d; see Methods). To determine the dynamics of this shift we varied the stimulation frequency. Only those experiments where the shift had occurred already with the second stimulus (Fig. 1f) were considered. The shift occurred for stimulation frequencies between 6.25 and 12.5 Hz ($n = 4$; Fig. 1h) and was due to an increase of

dendritic inhibition for interstimulus intervals (ISIs) of up to 300 ms, together with a decrease of somatic inhibition below the ISIs of 100 ms (Fig. 1g).

The smaller and slower recurrent IPSPs recorded in the dendrites with 0.2-Hz stimulation, or at the onset of a stimulus series, could reflect either an attenuated IPSP originating near the soma or the activation of a small and slow inhibitory conductance in the dendrites¹³. If the first hypothesis is correct, then preventing somatic hyperpolarization by voltage clamping (V-clamping) the soma should eliminate the IPSPs recorded in the dendrites. V-clamping the soma of CA1 pyramidal cells (V_m before V-clamp, $-58.0 \pm 0.8 \text{ mV}$; V_{holding} in V-clamp, $-58.2 \pm 0.7 \text{ mV}$; $n = 17$) while keeping the apical dendrite in the current-clamp (I-clamp) configuration (dendrite V_m , $-57.8 \pm 1.0 \text{ mV}$ before and $-57.4 \pm 0.9 \text{ mV}$ after V-clamping the soma; $n = 17$), strongly reduced the amplitude of the recurrent IPSP recorded in the dendrites when stimulating at 0.2 Hz ($64.5 \pm 4.1\%$ reduction; $P = 2.4 \times 10^{-6}$; $n = 17$; Fig. 2a, b).

This reduction was much larger than would be expected if IPSPs originated near the dendritic recording site. In fact, dynamically clamping (g-clamping) an inhibitory conductance (for g_{in} , $\tau_{\text{rise}} = 0.6 \text{ ms}$, $\tau_{\text{decay1}} = 3.9 \text{ ms}$ and $\tau_{\text{decay2}} = 14 \text{ ms}$; see Methods) with the dendritic recording pipette produced a simulated dendritic IPSP whose amplitude was attenuated, by V-clamping the soma, by only $13.5 \pm 2.7\%$ ($n = 4$; Fig. 2c, d). Dendritic inhibitory postsynaptic currents (IPSCs), however, may be slower than the one used for this simulation¹³ and, hence, be better V-clamped by the somatic pipette¹⁴. We therefore repeated the g-clamp experiment with slower conductances (Fig. 2d). Even a conductance eight times slower (for g_{in8} , $\tau_{\text{rise}} = 4.8 \text{ ms}$, $\tau_{\text{decay1}} = 31.2 \text{ ms}$ and $\tau_{\text{decay2}} = 112 \text{ ms}$) produced a simulated dendritic IPSP that was attenuated by only $38.5 \pm 2.5\%$ by V-clamping the soma.

These results indicate that recurrent IPSPs elicited at low stimulation frequencies are predominantly generated by perisomatic conductances. In contrast, IPSPs evoked with repetitive, high-frequency alveus stimulation (100 Hz) and recorded in the dendrites were much less affected by somatic V-clamp: $65.6 \pm 4.1\%$ and $15.5 \pm 7.0\%$ attenuation of the IPSP elicited with the 1st and 3rd stimulus respectively; $P = 2.9 \times 10^{-5}$; $n = 8$ (Fig. 2e, f). Furthermore, the amplitude of IPSCs recorded in the V-clamped soma progressively decreased with increasing stimulus number (to $51.1 \pm 15.6\%$ after the 3rd stimulus; $P = 0.038$, $n = 8$; Fig. 2g), consistent with the decline of the somatic IPSP slope (Fig. 1d).

These results demonstrate that after the onset of a series of stimuli delivered to the alveus, recurrent inhibitory conductances subside in the pyramidal cell layer and build up in more distal regions.

Interneurons responding to onset or frequency of spike series

Two types of recurrent circuit could account for the results above: (1) either CA1 pyramidal cells excite a population of interneurons that form synapses over extended portions of the somato-dendritic axis, with proximal synapses depressing and distal ones facilitating; or (2) pyramidal cells excite different types of interneurons, at the onset versus later in a series of spikes, that specifically target either proximal or distal compartments¹⁵. To distinguish between these two possibilities we recorded from interneurons located in the stratum oriens that are therefore likely to receive recurrent excitation from CA1 pyramidal cells. Recordings were performed in the loose-patch configuration to preserve the intracellular ionic composition, and the spiking activity of interneurons was monitored in response to series of ten stimuli at 50 Hz delivered to the alveus every five seconds (Fig. 3a). We gradually increased the stimulation intensity until the threshold for spike generation was reached within

the first four stimuli in the series (that is, until at least one of the first four stimuli triggered a spike in at least 20% of the trials, without exceeding 80%).

The spiking responses to this stimulation protocol allowed us to distinguish two populations of interneurons. In the first group (33 out of 49 cells), the probability of spiking was highest at the onset of the series and rapidly fell with subsequent stimuli ('onset-transient' interneurons; mean success rate for spike generation: $42.7 \pm 3.6\%$ after the first stimulus; $4.9 \pm 1.3\%$ after the third stimulus; Fig. 3a, d). In the second group (16 out of 49 cells) the probability of spike generation was lowest after the first stimulus and increased to a plateau between the third and tenth stimuli ('late-persistent' interneurons; mean success rate for spike generation: $1.6 \pm 1.6\%$ for the first stimulus; $28.9 \pm 6\%$ for the third stimulus; Fig. 3a, d). Spikes recorded in onset-transient interneurons showed very little variability in delay ('jitter'; standard deviation, $\pm 470 \mu\text{s}$; Fig. 3b), suggesting that they may operate as coincidence detectors^{8,16,17}. In contrast, spikes recorded in late-persistent interneurons were much less precise and jittered within a $\pm 2.11 \text{ ms}$ window (Fig. 3b).

Interneurons were then re-patched with an intracellular solution

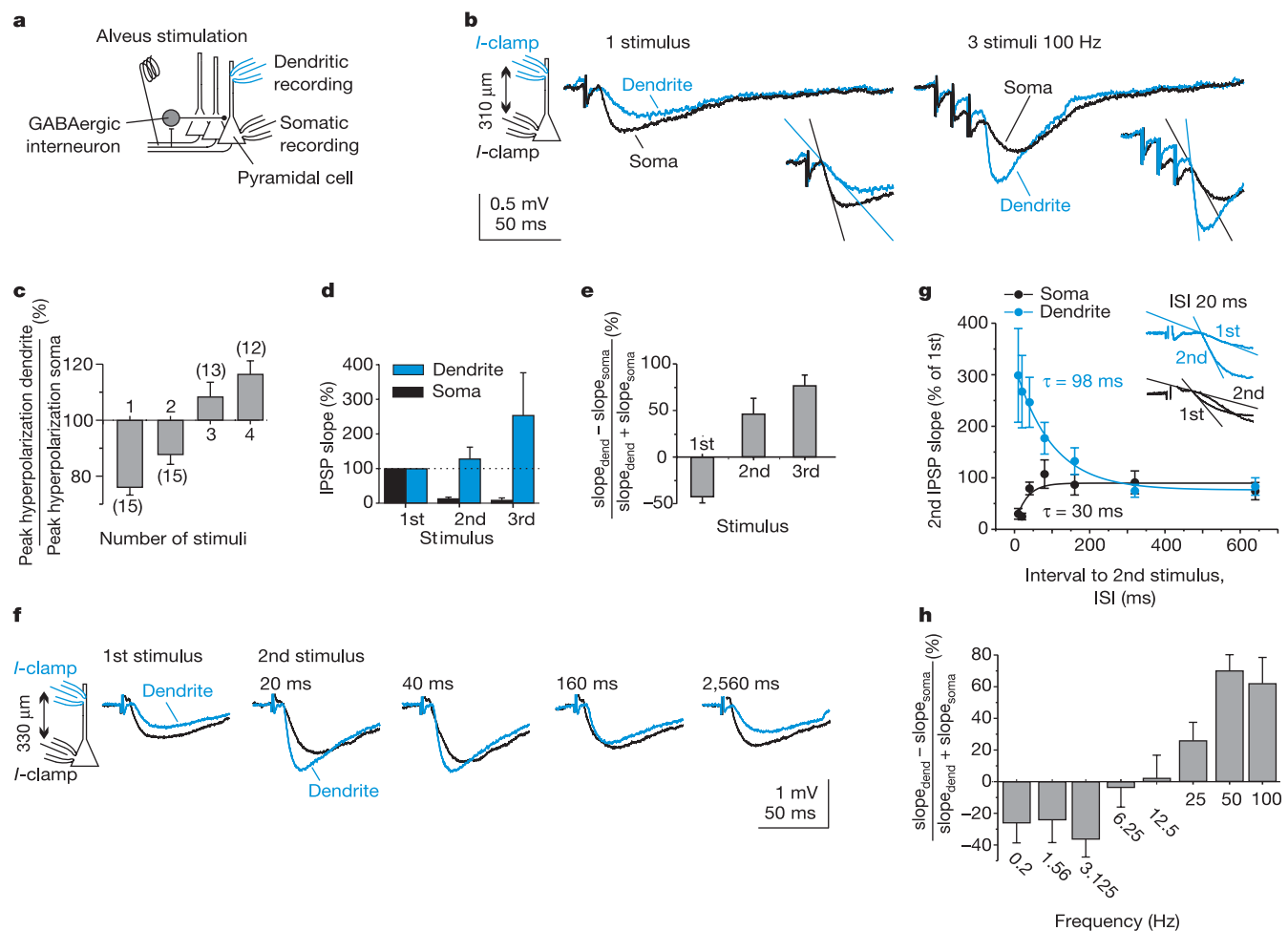


Figure 1 Shift of recurrent inhibition along the somato-dendritic axis. **a**, Experimental configuration. **b**, Somatic (black) and dendritic (blue) responses to 0.2 Hz (left) and 100 Hz (right) alveus stimulation. Insets, note shift in initial slope of the IPSPs. **c**, Ratio of dendritic versus somatic peak hyperpolarization plotted against the number of stimuli (n is shown in parentheses). **d**, IPSPs' slope plotted against stimulus number ($n = 15$; for isolation of individual IPSPs evoked by a given stimulus in a series, see Methods). **e**, Difference between dendritic and somatic IPSPs' slopes, normalized by their sum, and

plotted against stimulus number ($n = 15$). **f**, Somatic and dendritic IPSPs in response to pairs of stimuli with increasing interstimulus interval (ISI). **g**, Slope of the second IPSP plotted against ISI (single exponential fits, $n = 4$). Inset, despite the decrease in the initial slope of the second somatic IPSP, the increase in its peak amplitude results from the IPSP increase in the dendrite. **h**, As in **e** but plotted against stimulus frequency ($1/\text{ISI}$; $n = 4$).

containing biocytin 0.5% for morphological analysis. The radial distribution of their axons could be determined in 16 out of 49 cells (see Methods). The axons of onset transient interneurons (11 out of 16 recovered neurons) were distributed around the pyramidal cell layer. The average distribution peaked within the pyramidal cell layer and showed a half-width of about 100 μm (range 20–390 μm ; Fig. 3c, e). In contrast, the average axonal distribution of late-persistent interneurons (5 out of 16 recovered neurons) peaked 320 μm above the border between the stratum pyramidale and stratum radiatum (Fig. 3c, e) also with a half-width of about 100 μm . The axonal distribution of early-transient and late-persistent interneurons along the radial axis of the hippocampus is consistent with the axonal arborization of basket, axo-axonic, bistratified and trilaminar cells on the one hand and of oriens-lacunosum moleculare interneurons, on the other¹⁵. Future immunohistochemical analysis will allow us to determine the relative contribution of each interneuronal subtype contained in our two physiological categories. Nevertheless, the distribution of spike probabilities during a series of stimuli (Fig. 3d), together with the axonal arborization of early-transient and late-persistent interneurons (Fig. 3e), strongly suggest that the onset of recurrent inhibition in the pyramidal cell layer and its subsequent shift to more distal regions result from the recruitment of distinct interneuron types. Although it is likely that a single CA1 pyramidal cell axon may contact distinct types of local postsynaptic targets^{18,19}, as

has been shown for CA3 pyramidal cell axons^{20,21}, we cannot exclude the possibility that different interneuron types may be contacted by distinct CA1 pyramidal cells.

Coincidence detection and integration in different interneurons

Several properties that have been shown to differ between types of interneurons could account for the onset-transient versus late-persistent recruitment. These properties include the kinetics of excitatory postsynaptic currents^{16,22} (EPSCs), their short-term plastic properties^{18,19,23}, the membrane time constant²⁴ and the received synaptic inhibition^{25,26}. We therefore recorded synaptic activity evoked by alveus stimulation, in both onset-transient and late-persistent interneurons in the whole-cell *V*-clamp configuration (Fig. 4a–d). Initially, these experiments were performed in the presence of the GABA_A receptor antagonist gabazine (6–25 μM). The amplitude of EPSCs recorded in onset-transient interneurons depressed during the series of stimuli (to $32.4 \pm 4.8\%$ after the 10th stimulus; $n = 13$), whereas it facilitated in late-persistent interneurons (to $393.0 \pm 77.2\%$ after the 10th stimulus; $n = 9$) (Fig. 4a, b), consistent with previous observations^{18,19,23}. Recovery from facilitation of EPSCs recorded in late-persistent interneurons was slower (295 ms) than recovery from depression in onset-transient interneurons (11 ms; Fig. 4c), consistent with the slower recovery of dendritic than somatic IPSPs with increasing ISI, reported above (see Fig. 1g). Furthermore, onset-transient inter-

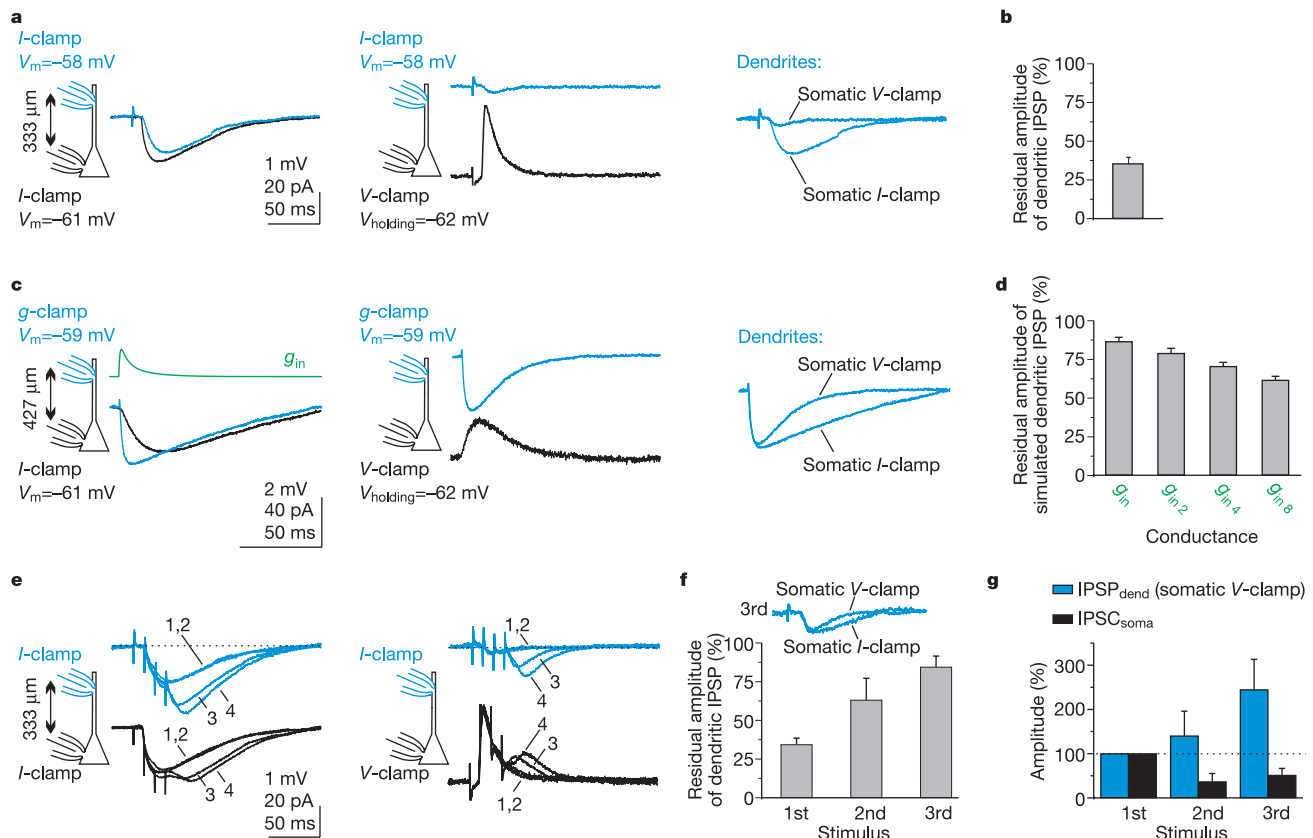


Figure 2 Transient activation of somatic and delayed activation of dendritic inhibitory conductances. **a**, Responses to 0.2-Hz alveus stimulation without (left) and with (centre) somatic *V*-clamp. Dendritic IPSPs are superimposed on the right for comparison. **b**, Residual amplitude of the dendritic IPSP after somatic *V*-clamp ($n = 17$). **c**, In a different cell, responses to dynamic current injection in the dendrite; conductance time course (g_{in}) in green, 5-nS peak amplitude. **d**, Residual amplitude of the simulated dendritic IPSPs after somatic *V*-clamp ($n = 4$) for conductance time courses as in **c**, or 2,

4 and 8 times longer. **e**, In the same cell as in **a**, responses to 1 stimulus, or 2, 3 and 4 stimuli at 100 Hz without (left) and with (centre) somatic *V*-clamp. Note that *V*-clamp abolishes dendritic IPSPs 1 and 2, but not 3 and 4. **f**, As in **b** but for IPSPs elicited by the 1st, 2nd and 3rd stimuli ($n = 8$; inset, superimposed third IPSPs isolated from traces in **e**). **g**, Peak amplitude of the dendritic IPSP (with somatic *V*-clamp) and of the somatic IPSC plotted against stimulus number ($n = 8$).

neurons showed EPSCs with significantly faster decay kinetics than late-persistent interneurons: single exponential fit, 1.77 ± 0.13 ms ($n = 11$) versus 4.11 ± 0.46 ms ($n = 7$); $P = 2.3 \times 10^{-5}$ (Fig. 4a, d).

To determine the impact that different EPSC kinetics have on the time course of the EPSP we imposed fast ($\tau_{\text{rise fast}} = 0.61$ ms, $\tau_{\text{decay fast}} = 1.77$ ms) and slow ($\tau_{\text{rise slow}} = 0.87$ ms, $\tau_{\text{decay slow}} = 4.11$ ms) excitatory conductances in the *g*-clamp mode in both onset-transient and late-persistent interneurons. The amplitudes of the fast and slow conductances were matched to produce artificial excitatory postsynaptic potentials (aEPSPs) with similar amplitudes (range 2–9 nS; Fig. 4e). The half-decay of the resulting aEPSPs were analysed separately for the two interneuron types because of the different membrane time constants (τ_m) displayed by these interneurons: early-transient, $\tau_m = 19.2 \pm 1.9$ ms ($n = 11$); late-transient, $\tau_m = 35.3 \pm 4.7$ ms ($n = 8$); $P = 0.0021$. In late-persistent interneurons, the half-decay of aEPSPs triggered by slow conductances was $24.2 \pm 5.9\%$ longer than when triggered by fast conductances (32.3 ± 5.4 ms versus 28.0 ± 5.9 ms, $P = 0.0023$; $n = 7$). The effect was even more pronounced in onset-transient interneurons, where slow conductances led to aEPSPs that were $53.5 \pm 5.2\%$ longer than those triggered by fast conductances (18.7 ± 2.4 ms versus 12.2 ± 1.5 ms, $P = 0.0036$; $n = 5$). Accordingly, the half-decay of EPSPs evoked with alveus stimulation was significantly shorter in onset-transient interneurons as compared to late-persistent interneurons (15.5 ± 2.7 ms versus 31.9 ± 3.4 ms, $P = 0.0031$; $n = 14$).

These data suggest that fast EPSC kinetics and membrane time constants may be essential in reducing temporal summation of

consecutive EPSPs in onset-transient interneurons¹⁶. In contrast, the slower EPSCs and membrane time constant observed in late-transient interneurons may prolong the time for effective summation of consecutive EPSPs. The resulting difference in temporal summation between the two interneurons is illustrated in Fig. 4f ($P = 0.023$; $n = 13$). In late-persistent interneurons the peak depolarization reached by the four last EPSPs in a series of ten stimuli is $1,265 \pm 308\%$ ($n = 6$) the amplitude of the first EPSP, that is, 3.1 times the increment due solely to synaptic facilitation. In contrast, the peak depolarization reached by the last EPSP in onset-transient interneurons is $59 \pm 6\%$ ($n = 7$) the amplitude of the first EPSP, that is, only 2.2 times larger than the value of synaptic depression.

We specifically addressed the role of EPSC kinetics on temporal summation by pharmacologically modifying their decay with CX546, a substance that slows deactivation and desensitization of AMPA receptors²⁷. We applied CX546 at a subsaturating concentration of 200–400 μ M, to prolong the decay of EPSCs recorded in onset-transient interneurons to a value that roughly matched the decay of EPSC recorded in late-persistent interneurons (before CX546, 2.25 ± 0.11 ms; after CX546, 4.44 ± 0.11 ms; $P = 44.2 \times 10^{-5}$; $n = 4$) without significantly affecting amplitude ($3.4 \pm 5.7\%$ decrease; $P = 0.59$; $n = 4$). CX546 did not significantly affect short-term depression ($P > 0.086$, for any stimulus in a train of ten stimuli at 12.5 Hz, $n = 4$; Fig. 4g), suggesting that EPSC depression (Fig. 4a–c) is not due to desensitization of AMPA receptors but instead, is mainly of presynaptic origin.

This prolongation in EPSC kinetics considerably increased the

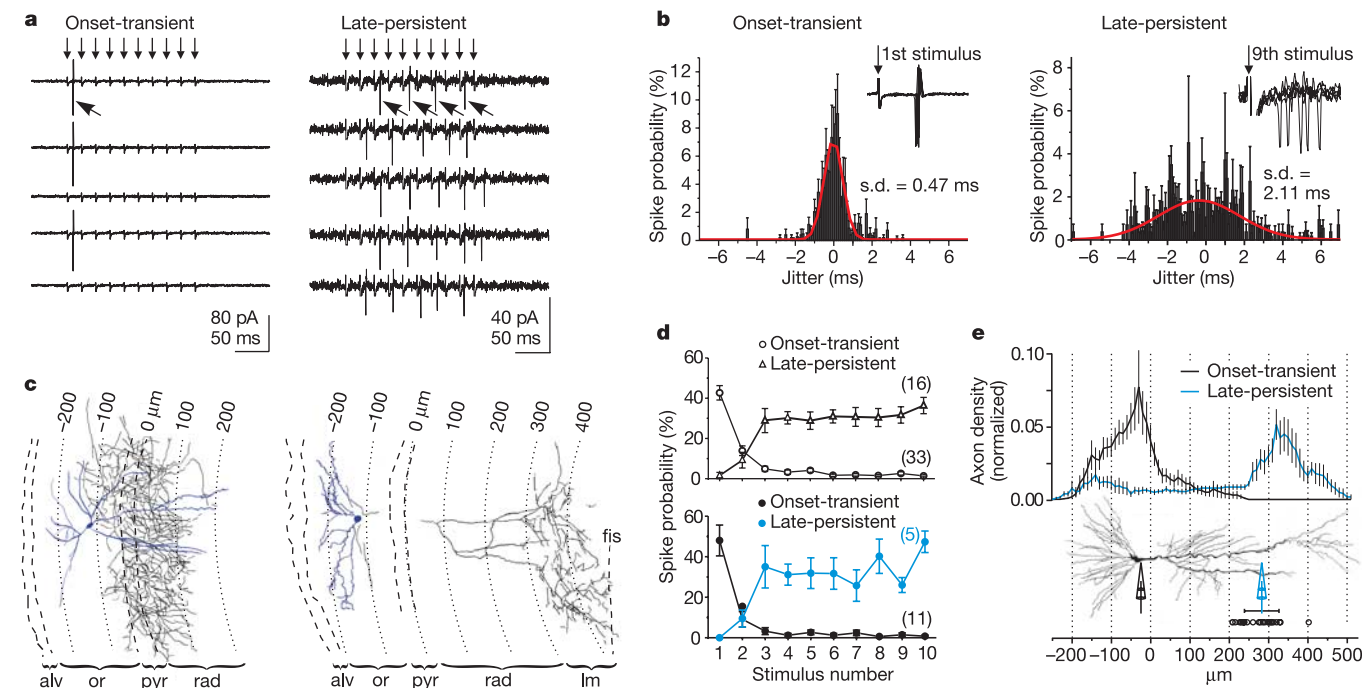


Figure 3 Onset-transient and late-persistent interneurons project to distinct layers. **a**, Loose-patch recording from 'onset-transient' (left) and 'late-persistent' (right) interneurons in the stratum oriens, in response to ten alveus stimulations (small arrows, 50 Hz) at threshold intensity. Large arrows, spikes. **b**, Jitter of spikes evoked by the first stimulus in onset-transient interneurons (left, $n = 30$) and by each of the last three stimuli in late-persistent interneurons (right, $n = 16$; bin width: 100 μ s; the abscissa shows the difference between the delay of each spike and their mean delay after a given stimulus; the gaussian fit is shown in red). Insets, five superimposed sweeps. **c**, Camera lucida reconstruction of the neurons recorded in **a** (dark blue, dendrite; black, axon). Alv, alveus;

or, stratum oriens; pyr, stratum pyramidal; rad, stratum radiatum; Im, stratum lacunosum moleculare; fis, fissure. **d**, Upper panel, spike probability plotted against stimulus number for onset-transient and late-persistent interneurons. Lower panel, same as above but restricted to anatomically recovered neurons. n is shown in parentheses. **e**, Axonal density for onset-transient and late-persistent interneurons plotted against radial distance from the border between strata pyramidal and radiatum (0- μ m line). A reconstructed CA1 pyramidal cell is aligned and scaled to the plot. Schematic recording pipettes indicate somatic and average dendritic ($\pm$ s.d., horizontal bar) recording sites for data illustrated in Figs 1 and 2 (open circles, the 28 dendritic recording sites).

summation of consecutive EPSPs in onset-transient interneurons: peak depolarization of the 10th EPSP normalized by the first EPSP; control, $32.7 \pm 10.3\%$; CX546, $100.8 \pm 12.7\%$, $P = 74.98 \times 10^{-4}$; ten stimuli at 50 Hz, $n = 4$ (Fig. 4g). Accordingly, CX546 also modified the spiking behaviour of onset-transient interneurons in response to series of ten stimuli, from a very transient to a more persistent one: when the first stimulus elicited a spike with a probability of $\sim 45\%$ in either condition, the average spiking probability between the 2nd and 10th stimulus decreased to $5.3 \pm 2.3\%$ in the control but remained at $31.7 \pm 4.5\%$ for CX546 ($P = 8.58 \times 10^{-8}$; $n = 4$; Fig. 4h). These results further demonstrate the crucial part played by EPSC kinetics in the integrative properties of onset-transient interneurons.

Finally, we determined the role played by disynaptic inhibition (see Methods) in the differential recruitment of interneurons by CA1 pyramidal cell activity^{25,26}. Alveus stimulation (in the absence of the GABA_A receptor antagonist gabazine) evoked EPSP–IPSP sequences in both interneuron types. The disynaptic IPSP decreased

the window for temporal summation of consecutive EPSPs (Fig. 5a, b). This effect was more marked at the beginning than at the end of a series of stimuli (Fig. 5a, b) because IPSCs depressed during repetitive stimulation (to $34.6 \pm 4.5\%$ at the 10th stimulus; $n = 10$; Fig. 5c). Hence, disynaptic IPSPs repolarize the membrane shortly after the first EPSP and are thus likely to contribute to the transience of the response in onset-transient interneurons. In late-persistent interneurons, on the other hand, the presence of disynaptic IPSPs delayed EPSP summation, thus reinforcing their late response. The inhibition/excitation ratio was computed as the gabazine-sensitive charge divided by the gabazine-resistant charge and was found to be 2.1 times larger in onset-transient than in late-persistent interneurons: 1.72 ± 0.16 ($n = 7$) versus 0.82 ± 0.55 ($n = 3$). (The gabazine-sensitive charge is the time integral of the control response minus the response in the presence of gabazine, labelled ‘subtraction’ in Fig. 5c; the gabazine-resistant charge is the time integral of the response in the presence of gabazine, labelled ‘gabazine’ in Fig. 5c.) These results clearly indicate that disynaptic

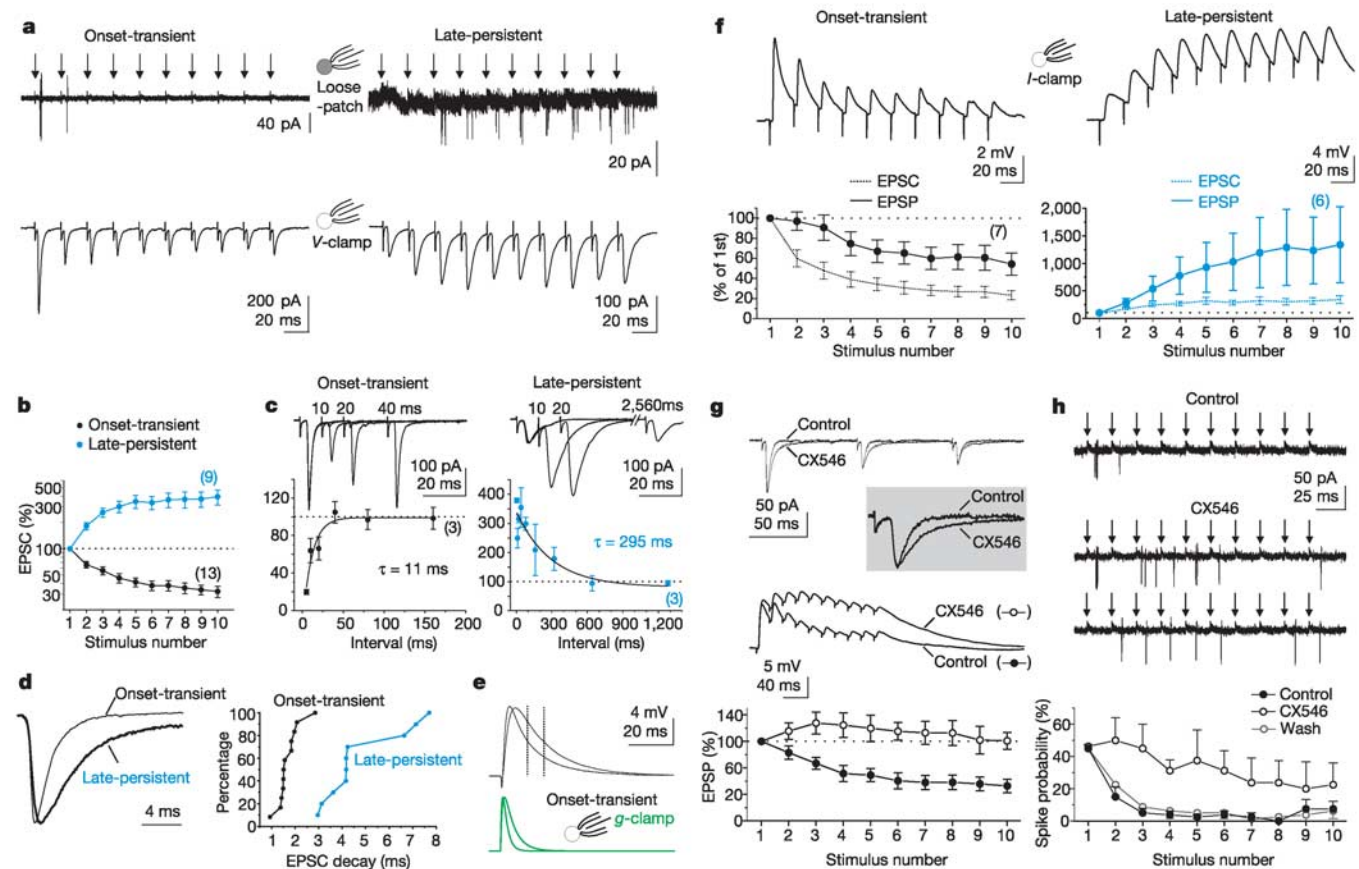


Figure 4 Coincidence detection and integration in onset-transient versus late-persistent interneurons. **a**, Recordings from an onset-transient (left) and a late-persistent (right) interneuron: responses to ten threshold alveus stimulations (50 Hz). Top panel, loose-patch, six superimposed sweeps (no gabazine). Bottom panel, whole-cell, V-clamp (gabazine present, 12.5 μ M). **b**, EPSCs' amplitude plotted against stimulus number (50 Hz; n shown in parentheses). **c**, EPSCs elicited by paired stimuli with increasing ISI. Summary graph: recovery from short-term depression in onset-transient interneurons and from short-term facilitation in late-persistent interneurons (single exponential fits). **d**, Peak scaled EPSCs from the two interneurons illustrated in **a**. Summary graphs: cumulative distribution of EPSCs' decay time constants. **e**, aEPSPs elicited by *g*-clamping a fast and a slow conductance (green; 9-nS and 6-nS peak amplitude, respectively) in an onset-transient interneuron. Dotted lines, half-decays. **f**, Same interneurons as in **a** but in *I*-clamp (stimulus intensity as in *V*-clamp). The summary graphs show the EPSPs' maximal

depolarization and EPSCs' amplitude plotted against stimulus number. **g**, Onset-transient interneuron with or without CX546 (200 μ M; gabazine present). Note that CX546 prolongs the decay of the EPSC (inset, expanded timescale) without affecting short-term depression (top panel). Bottom panel, same interneuron in *I*-clamp (50-Hz stimulation). The summary graph shows maximal depolarization plotted against stimulus number ($n = 4$). **h**, In a different onset-transient interneuron, loose-patch recordings with or without CX546 (200 μ M; no gabazine). Top panel, six superimposed sweeps (50% spike probability after stimulus one). Middle and bottom panels, three superimposed sweeps per panel, ordered depending on whether stimulus one elicited a spike or not (50% spike probability after stimulus one was maintained in CX546 by decreasing stimulation intensity). The summary graph shows spike probability plotted against stimulus number ($n = 4$).

inhibition contributes to temporally segregating the activation of early-transient versus late-persistent interneurons by CA1 pyramidal cells.

Selective routing via two neuronal modules

This circuit analysis reveals the interplay between four properties, namely: EPSC kinetics, the membrane time constant, short-term plasticity and disynaptic inhibition of interneurons. These properties are synergistically tuned to produce recurrent inhibitory circuits that operate in two distinct modes: coincidence detection and integration. The two modes extract two different temporal features of a spike series and route them as independent signals to different hippocampal layers. The coincidence-detection mode reports the onset of a spike series by activating perisomatic inhibitory conductances on pyramidal cells. The high frequency firing of basket cells observed *in vivo*²⁸ may, hence, signal the sequential onset of activity of independent excitatory pathways converging onto them. The integration mode, on the other hand, activates distal dendritic inhibitory conductances in proportion to the rate of the spikes in a series (Fig. 5d).

We therefore expect the wide range of temporal activity patterns characteristic of CA1 pyramidal cells during exploratory behaviour in rats²⁹ to be represented as a spatially distinct inhibitory pattern across hippocampal layers. The temporal precision and perisomatic location of the coincidence-detector module may help synchronize CA1 pyramidal cell activity with the onset of a spike series.

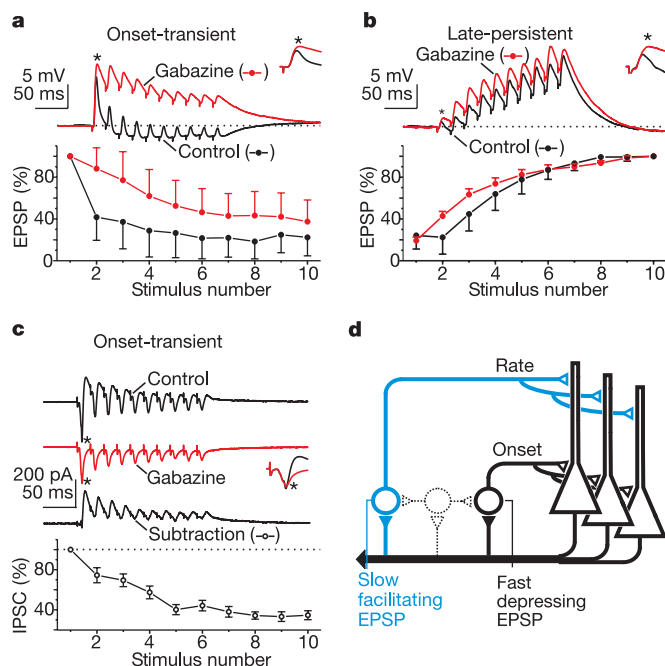


Figure 5 Disynaptic inhibition contributes to transience in onset-transient and to delay in late-persistent interneurons. **a**, Onset-transient interneuron: *I*-clamp responses to ten stimuli (50 Hz) with or without gabazine (12.5 μM). Inset, note the lack of effect of gabazine on the initial slope of the first EPSP (asterisk), indicating that inhibition is disynaptic. Summary graph: maximal depolarization of EPSPs plotted against stimulus number ($n = 4$). **b**, Late-persistent interneuron, stimulation and recording as in **a** (summary graph: $n = 4$). **c**, Onset-transient interneuron: *V*-clamp responses with or without gabazine and their algebraic difference to isolate disynaptic IPSCs. The summary graph shows the amplitude of disynaptic IPSCs plotted against stimulus number (onset-transient: $n = 7$; late-persistent: $n = 3$). **d**, Schematic illustration of recurrent inhibition on pyramidal cells: onset-transient (black) and late-persistent (blue) loops (the hypothetical GABAergic interneuron represented in dotted lines might be identical to an onset-transient interneuron inhibiting itself and late-persistent interneurons).

Subsequent spikes in the series will affect more distal layers, thereby specifically modulating distal excitatory inputs. Interestingly, inhibitory interneurons with similar axonal projections and synaptic properties have also been described in the neocortex^{30,31}. Future work will elucidate whether a frequency-dependent shift of recurrent inhibition along the somato-dendritic axis is an attribute common to other cortical circuits. □

Methods

Slice preparation and electrophysiology

Hippocampal slices (400 μm) were prepared from 28- to 32-day-old Wistar rats as previously described⁸. Artificial cerebrospinal fluid (ACSF) consisted of (in mM): 119 NaCl, 2.5 KCl, 1.3 NaHPO₄, 1.3 MgCl₂, 2.5 CaCl₂, 26 NaHCO₃, 11 glucose (equilibrated with 95% O₂ and 5% CO₂). Experiments were performed at 31–32 °C, in the presence of the GABA_B receptor antagonist CGP62349 (1–3 μM) and the NMDA receptor antagonist D-CPP (25 μM). Whole-cell recordings were performed with patch pipettes (3–7 MΩ for somatic and 13–18 MΩ for dendritic recordings) filled with (in mM): 150 K-gluconate, 0.5 MgCl₂, 5 HEPES, 1.1 EGTA, and 10 phosphocreatine (pH = 7.25; 280–290 mOsm); biocytin 0.5% (for interneuron recordings). Loose-patch recording pipettes (10–15 MΩ) contained ACSF. Stimulation (100 μs) was applied using monopolar steel electrodes or ACSF-containing glass pipettes (3–7 MΩ). Voltage measurements were not corrected for the experimentally determined junction potential ($V_{ACSF} - V_{internal\ solution} = -12 \pm 0.4$ mV; $n = 3$). Data were recorded using Multiclamp 700A, Axopatch 200B and Axoclamp 2B amplifiers (digitalization at 5–10 kHz). The drugs used were NBQX, D-CPP, SR95531 ('gabazine'), CGP62349 and CX546. All experiments were carried out in accordance with the guidelines set out by the University of California, San Diego and the University of Zürich, Switzerland.

Alveus stimulation

The high Cl[−] driving force (~20 mV) allowed the detection of IPSP/IPSCs with small underlying conductances in CA1 pyramidal cells and hence the use of weak stimulation intensities: $V_{holding}$: 57.1 ± 2.1 mV; IPSCs peak amplitude, 28.9 ± 6.6 pA; $V_{reversal}$: 78.9 ± 3.2 mV; peak conductance, 1.34 ± 0.22 nS; $n = 6$. The IPSCs were abolished ($99.7 \pm 0.1\%$) by the selective GABA_A-receptor antagonist gabazine (3–10 μM; $n = 9$). IPSCs recorded in CA1 pyramidal cells were disynaptic as they were blocked by the AMPA/kainate receptor antagonist NBQX (20–25 μM; $96.0 \pm 1.3\%$ reduction; delay between the stimulus and the onset of the IPSC: 6.2 ± 0.3 ms; $n = 20$; Supplementary Fig. 1). IPSCs recorded in interneurons were disynaptic as the initial slope of the EPSC/P did not significantly change after gabazine perfusion⁸ ($3.1 \pm 2\%$ increase, $P = 0.82$, $n = 13$; Fig. 5). IPSP/IPSCs were not contaminated by Schaffer collateral stimulation (feed-forward IPSP/IPSC) or recurrent excitation between CA1 pyramidal cells (multisynaptic IPSP/IPSC): see Supplementary Information.

Analysis

Average values in the text and the Figures are expressed as mean $\pm$ s.e.m. The Student *t*-test was used for statistical comparisons. All traces are averages of 10 to 40 sweeps, unless stated otherwise. During a series of stimuli, individual somatic and dendritic IPSP/IPSCs were isolated by subtracting the averaged response to $n - 1$ stimuli from the one to n stimuli (Figs 1d–h and 2e–g). For slope measurements, the IPSP with the steepest initial slope was selected (usually the 1st for somatic recordings and the 2nd or 3rd or 4th for dendritic recordings) and used to define the size and the position of a time window (1–2 ms) to measure the slope of all IPSPs in the series (Fig. 1d–h). Membrane time constants were estimated by fitting a single exponential to the late portion of the membrane potential relaxation to a step current injection of −25 to −50 pA.

Dynamic clamp

The operation was performed as described previously⁸. Artificial inhibitory conductances (5–nS peak amplitude, $E_{rev} = -85$ mV) were imposed with the dendritic recording pipette only when the dendritic series resistance was ≤ 30 MΩ and could be properly balanced. The four kinetics used were: for g_{in1} , $\tau_{rise} = 0.6$ ms, $\tau_{decay1} = 3.9$ ms, $\tau_{decay2} = 14$ ms; for g_{in2} , $\tau_{rise} = 1.2$ ms, $\tau_{decay1} = 7.8$ ms, $\tau_{decay2} = 28$ ms; for g_{in3} , $\tau_{rise} = 2.4$ ms, $\tau_{decay1} = 15.6$ ms, $\tau_{decay2} = 56$ ms; for g_{in4} , $\tau_{rise} = 4.8$ ms, $\tau_{decay1} = 31.2$ ms, $\tau_{decay2} = 112$ ms. Ratio of the amplitudes of the two decay components: 0.634/0.366. Artificial excitatory conductances imposed in interneurons (2–9 nS peak amplitude, $E_{rev} = 0$ mV) had two different kinetics: for fast conductances, $\tau_{rise\ fast} = 0.61$ ms, $\tau_{decay\ fast} = 1.77$ ms; for slow conductances, $\tau_{rise\ slow} = 0.87$ ms, $\tau_{decay\ slow} = 4.11$ ms.

Morphology

Anatomical recovery of biocytin-filled interneurons: see Supplementary Methods. Interneuron soma, axons and dendrites were reconstructed using a drawing table (Olympus Optical). To quantify the axonal density along the radial axis, we counted the number of intersections made by the axons with lines running parallel (or in some cases concentric) to the border between the pyramidal cell layer and the stratum radiatum and interspaced by 10 μm.

Received 27 February; accepted 7 May 2004; doi:10.1038/nature02615.

Published online 30 May 2004.

- Adrian, E. D. *The Basis of Sensation: The Action of the Sense Organs* (W. W. Norton, New York, 1928).
- Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. *Spikes: Exploring the Neural Code*

- (The MIT Press, Cambridge, Massachusetts/London, UK, 1997).
3. Dayan, P. & Abbott, L. F. *Theoretical Neurosciences: Computational and Mathematical Modeling of Neural Systems* (The MIT Press, Cambridge, Massachusetts/London, UK, 2001).
 4. Andersen, P., Eccles, J. C. & Loynning, Y. Recurrent inhibition in the hippocampus with identification of the inhibitory cell and its synapses. *Nature* **198**, 540–542 (1963).
 5. Megias, M., Emri, Z., Freund, T. F. & Gulyás, A. I. Total number and distribution of inhibitory and excitatory synapses on hippocampal CA1 pyramidal cells. *Neuroscience* **102**, 527–540 (2001).
 6. Stuart, G. J., Dodt, H. U. & Sakmann, B. Patch-clamp recordings from the soma and dendrites of neurons in brain slices using infrared video microscopy. *Pflügers Arch.* **423**, 511–518 (1993).
 7. Spruston, N., Schiller, Y., Stuart, G. & Sakmann, B. Activity-dependent action potential invasion and calcium influx into hippocampal CA1 dendrites. *Science* **268**, 297–300 (1995).
 8. Pouille, F. & Scanziani, M. Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science* **293**, 1159–1160 (2001).
 9. Kandel, E. R., Spencer, W. A. & Brinley, F. J. Electrophysiology of hippocampal neurons. I. Sequential invasion and synaptic organization. *J. Neurophysiol.* **24**, 225–242 (1961).
 10. Dingledine, R. & Langmoen, I. A. Conductance changes and inhibitory actions of hippocampal recurrent IPSPs. *Brain Res.* **185**, 277–287 (1980).
 11. Alger, B. E. & Nicoll, R. A. Feed-forward dendritic inhibition in rat hippocampal pyramidal cells studied *in vitro*. *J. Physiol.* **328**, 105–123 (1982).
 12. Maccaferri, G. & McBain, C. J. Passive propagation of LTD to stratum oriens-alveus inhibitory neurons modulates the temporoammonic input to the hippocampal CA1 region. *Neuron* **15**, 137–145 (1995).
 13. Pearce, R. A. Physiological evidence for two distinct GABA_A responses in rat hippocampus. *Neuron* **10**, 189–200 (1993).
 14. Spruston, N., Jaffe, D. B., Williams, S. H. & Johnston, D. Voltage- and space-clamp errors associated with the measurement of electrotonically remote synaptic events. *J. Neurophysiol.* **70**, 781–802 (1993).
 15. Freund, T. F. & Buzsáki, G. Interneurons of the hippocampus. *Hippocampus* **6**, 347–470 (1996).
 16. Geiger, J. R. P., Lubke, J., Roth, A., Frotscher, M. & Jonas, P. Submillisecond AMPA receptor-mediated signaling at a principal neuron-interneuron synapse. *Neuron* **18**, 1009–1023 (1997).
 17. Fricker, D. & Miles, R. EPSP amplification and the precision of spike timing in hippocampal neurons. *Neuron* **28**, 559–569 (2000).
 18. Ali, A. B. & Thomson, A. M. Facilitating pyramid to horizontal oriens-alveus interneurone inputs: dual intracellular recordings in slices of rat hippocampus. *J. Physiol.* **507.1**, 185–199 (1998).
 19. Ali, A. B., Deuchars, J., Pawelzik, H. & Thomson, A. M. CA1 pyramidal to basket and bistratified cell EPSPs: dual intracellular recordings in rat hippocampal slices. *J. Physiol.* **507.1**, 201–219 (1998).
 20. Shigemoto, R. *et al.* Target-cell-specific concentration of a metabotropic glutamate receptor in the presynaptic active zone. *Nature* **381**, 523–525 (1996).
 21. Scanziani, M., Gähwiler, B. H. & Chazpak, S. Target cell-specific modulation of transmitter release at terminals from a single axon. *Proc. Natl Acad. Sci. USA* **95**, 12004–12009 (1998).
 22. Toth, K., Suares, G., Lawrence, J. J., Philips-Tansey, E. & McBain, C. J. Differential mechanisms of transmission at three types of mossy fiber synapse. *J. Neurosci.* **20**, 8279–8289 (2000).
 23. Losonczy, A., Zhang, L., Shigemoto, R., Somogyi, P. & Nusser, Z. Cell type dependence and variability in the short-term plasticity of EPSCs in identified mouse hippocampal interneurons. *J. Physiol.* **542**, 193–210 (2002).
 24. Buhl, E. H., Szilágyi, T., Halasy, K. & Somogyi, P. Physiological properties of anatomically identified basket and bistratified cells in the CA1 area of the rat hippocampus *in vitro*. *Hippocampus* **6**, 294–305 (1996).
 25. Gulyás, A. I., Megias, M., Emri, Z. & Freund, T. F. Total number and ratio of excitatory and inhibitory synapses converging onto single interneurons of different types in the CA1 area of the rat hippocampus. *J. Neurosci.* **19**, 10082–10097 (1999).
 26. Banks, M. I., White, J. A. & Pearce, R. A. Interactions between distinct GABA(A) circuits in hippocampus. *Neuron* **25**, 449–457 (2000).
 27. Arai, A. C., Xia, Y.-F., Rogers, G., Lynch, G. & Kessler, M. Benzamide-type AMPA receptor modulators form two subfamilies with distinct modes of action. *J. Pharmacol. Exp. Ther.* **303**, 1075–1085 (2002).
 28. Klausberger, T. *et al.* Brain-state- and cell-type-specific firing of hippocampal interneurons *in vivo*. *Nature* **421**, 844–849 (2003).
 29. O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
 30. Somogyi, P., Tamas, G., Lujan, R. & Buhl, E. H. Salient features of synaptic organisation in the cerebral cortex. *Brain Res. Brain Res. Rev.* **26**, 113–135 (1998).
 31. Beierlein, M., Gibson, J. R. & Connors, B. W. Two dynamically distinct inhibitory networks in layer 4 of the neocortex. *J. Neurophysiol.* **90**, 2987–3000 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Gähwiler from the Brain Research Institute of the University of Zürich, where the initial part of the study was performed. We thank J. Anderson for instructions on the use of the camera lucida. We also thank U. Gerber and M. Carandini for critical reading of and L. Glickfeld for comments on the manuscript. This work was supported by the Swiss National Science Foundation and NIH grants.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S. (massimo@ucsd.edu).

Structural analysis of a eukaryotic sliding DNA clamp–clamp loader complex

Gregory D. Bowman^{1,2}, Mike O'Donnell³ & John Kuriyan^{1,2}

¹Howard Hughes Medical Institute, Department of Molecular and Cell Biology and Department of Chemistry, University of California, Berkeley, California 94720, USA

²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

³Laboratory of DNA Replication, Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10021, USA

Sliding clamps are ring-shaped proteins that encircle DNA and confer high processivity on DNA polymerases. Here we report the crystal structure of the five-protein clamp loader complex (replication factor-C, RFC) of the yeast *Saccharomyces cerevisiae*, bound to the sliding clamp (proliferating cell nuclear antigen, PCNA). Tight interfacial coordination of the ATP analogue ATP- γ S by RFC results in a spiral arrangement of the ATPase domains of the clamp loader above the PCNA ring. Placement of a model for primed DNA within the central hole of PCNA reveals a striking correspondence between the RFC spiral and the grooves of the DNA double helix. This model, in which the clamp loader complex locks onto primed DNA in a screw-cap-like arrangement, provides a simple explanation for the process by which the engagement of primer–template junctions by the RFC:PCNA complex results in ATP hydrolysis and release of the sliding clamp on DNA.

Sliding clamps are DNA-tracking platforms that are essential for processive DNA replication in all living organisms^{1–5}. Sliding clamps enable DNA polymerases to cope with the considerable torque that results from the production of double-helical DNA, by allowing them to relax and regain their hold on DNA without losing their place at the replication fork. The efficient movement of the replication fork also relies critically on the rapid placement of sliding clamps at newly primed sites on the lagging DNA strand by ATP-dependent clamp loader complexes. This means that the piecewise generation of Okazaki fragments can keep up with the continuous synthesis of DNA on the leading strand. In addition to their role in DNA replication, sliding clamps are also involved in several other processes that require a mobile contact on DNA⁶.

The eukaryotic clamp loader assembly, RFC, consists of five subunits and forms a stable ATP-dependent complex with the eukaryotic sliding clamp, PCNA^{7–10}. The RFC:PCNA complex binds specifically to primed DNA, and the recognition of the double-stranded/single-stranded junction stimulates ATP hydrolysis by the clamp loader. This results in the dissociation of RFC from the clamp, leaving PCNA encircling DNA and ready to receive DNA polymerase for the next round of DNA synthesis (Fig. 1a).

The general architecture of the pentameric clamp loader assembly has been revealed by the crystal structure of an inactive form of the *Escherichia coli* clamp loader complex¹¹, and the structure of one of the subunits of the archaeobacterial clamp loader bound to ADP¹². Each clamp loader subunit contains two structurally conserved domains that together comprise an ATPase module of the AAA+ family^{13,14}. AAA+ ATPases are a diverse class of oligomeric (typically hexameric) proteins that couple ATP binding and hydrolysis to the modulation of protein–protein interactions¹³. Many AAA+ ATPases form ring-like assemblies, in which a characteristic feature is the presentation by each subunit of an arginine residue, termed the ‘arginine finger’, towards ATP bound at the next subunit^{13,15,16}.

In contrast to the symmetrical organization of the hexameric rings formed by AAA+ proteins such as N-ethylmaleimide-sensitive factor (NSF)^{15,16}, the structure of the inactive bacterial clamp loader complex is strikingly asymmetric¹¹, and it has been difficult to infer the molecular details of clamp loader action from this structure¹⁷. One important unresolved issue concerns the

mechanism by which ATP allows the specific recognition of primed DNA by the clamp loader. Also unknown is the nature of the ‘ejection mechanism’ by which binding to primed DNA results in the hydrolysis of ATP so that the complex is disassembled, the clamp is loaded onto DNA, and the clamp loader is released for another cycle of clamp loading^{10,18}.

We now describe the crystal structure of the *S. cerevisiae* RFC complex bound to the PCNA clamp. The structure reveals an unexpected ATP-dependent spiral arrangement of the ATPase domains of RFC that extends above the central channel of PCNA. The geometry of the spiral is such that the ATPase domains track the grooves of primed DNA modelled within the hole of the sliding clamp. The crystal structure suggests that DNA binding would stabilize a catalytically competent configuration of the interfacial ATP-binding sites, thereby providing a hypothesis for how the interaction with the double helix triggers the hydrolysis of ATP.

Structure determination

We have crystallized a variant form of the *S. cerevisiae* RFC complex in which the largest subunit (RFC-A in the notation introduced in Fig. 1) is truncated to a minimal construct (residues 295–785) that is fully functional in clamp loading (data not shown). Because ATP hydrolysis weakens the interaction between clamp and clamp loader, we replaced the arginine finger residues that are present in the highly conserved Ser-Arg-Cys (SRC) motifs of four RFC subunits (RFC-B, RFC-C, RFC-D and RFC-E) by glutamine, which resulted in a significant reduction in ATPase activity (Supplementary Fig. 1). Crystals of this mutated RFC:PCNA complex obtained in the presence of ATP- γ S diffract X-rays to only ~5 Å resolution using synchrotron radiation. Substantial improvement in crystal quality was obtained by controlled dehydration of the crystals¹⁹, allowing the structure to be determined to a resolution of 2.85 Å (Supplementary Table 1).

PCNA-bound RFC forms an ATP-dependent spiral

The hetero-pentameric RFC complex is seated on top of a closed PCNA ring, but tipped away from it (Fig. 1b). Domain I of each RFC subunit is a RecA-type ATPase domain, which is followed by a small helical domain (domain II) that is characteristic of AAA+ ATPases.

Domains I and II together form the AAA+ module, which is connected by a flexible linker to another helical domain (domain III) that is unique to clamp loaders (Fig. 1c, d). The helical third domains of all five RFC subunits pack together to form a stable cylindrical structure, referred to as the 'collar'¹¹. The five AAA+ modules of RFC assemble into a right-handed spiral, which results in only three of the five RFC subunits (RFC-A, RFC-B and RFC-C) making contact with PCNA, leaving a wedge-shaped gap between RFC-E and PCNA (Fig. 2).

ATP- γ S is bound to each of RFC-A, RFC-B, RFC-C and RFC-D (Supplementary Fig. 2). These nucleotides are key elements that hold the spiral assembly together by anchoring intersubunit interactions through hydrogen bonds to the phosphate groups (Fig. 3b). RFC-E lacks several residues that are conserved at the active sites of functional ATPases, but electron density maps reveal a strongly bound nucleotide (either ADP or ATP- γ S) at the RFC-E binding site (Supplementary Fig. 2). This nucleotide packs closely against a unique carboxy-terminal domain (domain IV) of RFC-A, which is situated between the ATPase domains of RFC-A and RFC-E, and provides a physical link between the two ends of the RFC spiral.

PCNA has three conserved hydrophobic grooves for potential interaction with RFC, and two of these sites are engaged by RFC-A and RFC-C. RFC-B, located between RFC-A and RFC-C, makes limited and primarily polar interactions with the intersubunit region of the clamp, and may not contribute significantly to binding energy. All three RFC subunits interact with PCNA primarily through the C-terminal end of the clamp-interacting helix (α 4; Fig. 3a) and the loop following it. The interface between RFC-A and PCNA resembles that of other proteins bound to sliding clamps^{20,21}, and involves a 3₁₀ helical turn that follows helix α 4 of RFC-A and inserts four hydrophobic residues (A-Val401, A-Val402, A-Tyr404 and A-Phe405) into the hydrophobic binding site on the clamp. Two hydrophobic residues of RFC-C (C-Ile106 and C-Phe107) interact with the corresponding site on the next subunit of the clamp.

The structure of PCNA in the RFC:PCNA complex is not distorted significantly with respect to the structure of free PCNA⁵, and we conclude that the crystal structure represents a stable interaction of the RFC complex with a closed PCNA clamp. In the clamp loading cycle, a clamp loader would be bound to a closed clamp immediately before ATP hydrolysis and dissociation of the

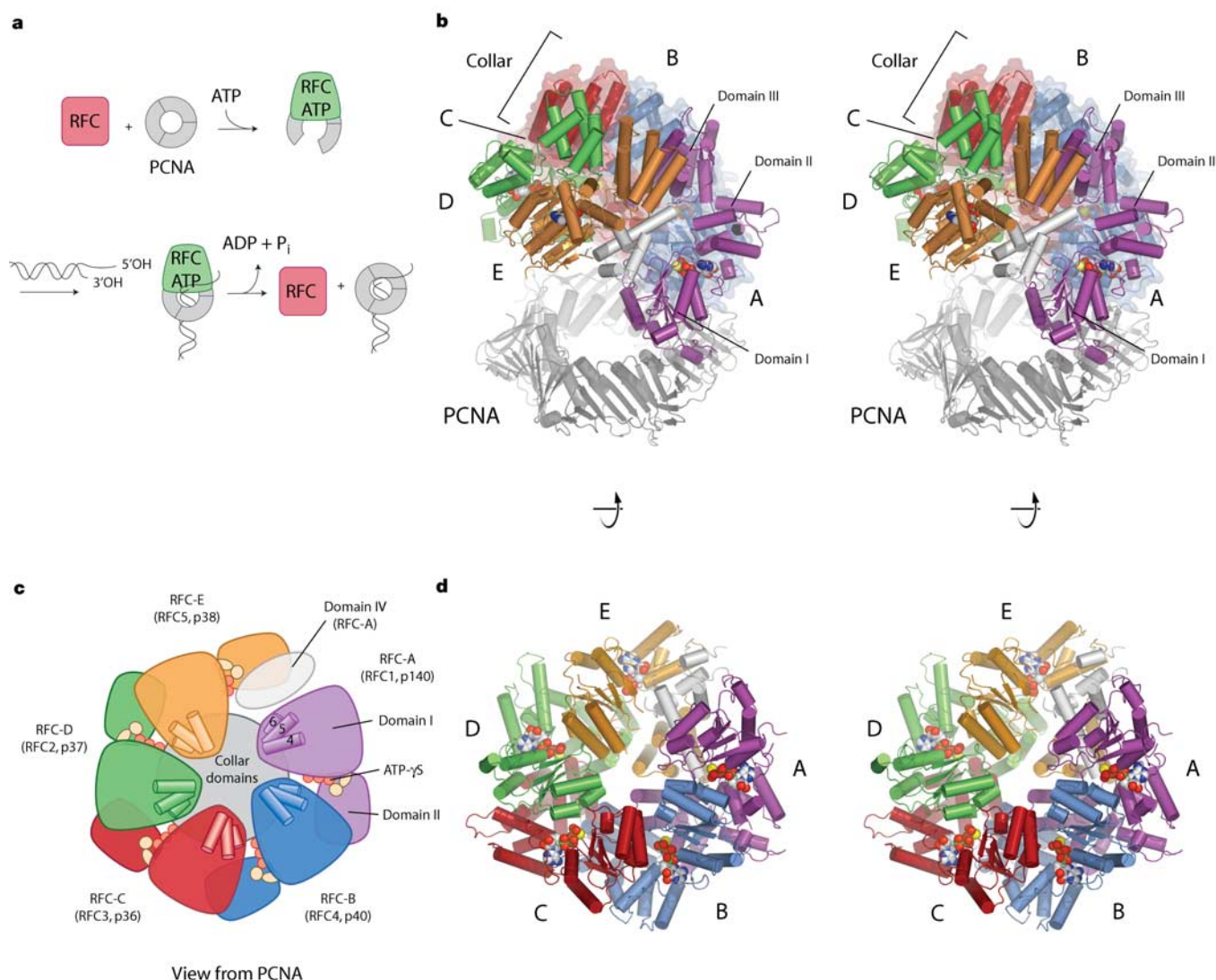


Figure 1 Overview of the RFC:PCNA complex. **a**, The clamp loading cycle. **b**, A stereoview of the RFC:PCNA complex, viewed from the side. **c**, **d**, Schematic (**c**) and backbone (**d**) representations of RFC rotated $\sim 110^\circ$ about the horizontal with respect to **b**. Viewed from the PCNA-interacting face of the complex, the five AAA+ modules are in the foreground,

with the C-terminal collar in the rear. The five subunits of the RFC complex are referred to as RFC-A, RFC-B, RFC-C, RFC-D and RFC-E, respectively, moving in a right-handed sense around the assembly, with the thumb pointing towards the collar. The yeast and human nomenclature for each subunit is shown in parentheses.

NATURE | VOL 429 | 17 JUNE 2004 | www.nature.com/nature

to generate a model for an RFC:DNA complex based purely on geometrical considerations. The axis of an ideal B-form DNA duplex was made coincident with the screw axis that relates RFC-A to RFC-E. The sequence-independent recognition of DNA is expected to occur in the minor groove, and so we rotated the DNA about its axis so that RFC-A was positioned alongside the minor groove. Each of the subunits of the RFC complex in this model is seen to track the minor groove of the double helix (Fig. 4a), with a fit that is tight but reasonable. This means that the architecture of the RFC:PCNA complex matches that of DNA not just in terms of the screw rotational operators, but also in terms of preserving the size of the internal chamber, so that it fits tightly around the double helix while allowing the preceding segment of DNA to be positioned comfortably within the PCNA ring (Fig. 4b).

The recognition of RNA–DNA hybrid structures is critical for the function of other clamp loader complexes, such as the *E. coli* γ -complex. Notably, the helical pitch of A-form RNA–DNA duplexes (5.5 Å per 60° turn) closely matches that of B-form DNA–DNA duplexes (5.6 Å per 60° turn)²⁸, which means that the distinction between the type of primer structure recognized by the clamp loader is not a crucial aspect of the geometrical analysis presented here.

Proposed DNA interaction

The amino-terminal regions of three α -helices of domain I of each RFC subunit ($\alpha 4$, $\alpha 5$, $\alpha 6$; the distal helices, see Fig. 3a), as well as the loops leading into them, are positioned so that they interact with the backbone of the modelled double helix (Fig. 4). The helix dipole effect from the central ($\alpha 5$) and clamp-interacting ($\alpha 4$) helices of each subunit generates a track of positive electrostatic potential within the inner surface of the spiral. In addition, each RFC subunit presents several lysine and arginine residues in a manner that suggests potential interaction with the DNA double helix (Fig. 5). These residues are conserved in all five RFC subunits so that, like the helix dipoles, they track the minor groove of the double helix.

The α -helices and the residues that are potentially important for DNA interaction are conserved in clamp loader subunits from eukaryotes, prokaryotes and archaea (Fig. 5). For example, A-Lys462 in RFC-A is in the loop preceding the SRC helix ($\alpha 6$), and it points directly towards the phosphate backbone of modelled DNA. This residue is either lysine or arginine in all eukaryotic and archaeal RFC sequences, as well as in the sequences of the γ -subunits of bacterial clamp loaders. Likewise, A-Arg383 in RFC-A is invariant in all eukaryotic and archaeal RFC sequences, and is conserved as Arg, Lys or His in 55% of bacterial sequences (bacterial clamp

loaders have conserved basic residues at nearby positions that are not conserved in the RFC sequences). A third highly conserved position lies in the loop preceding the central helix ($\alpha 5$). A-Arg434 is conserved as arginine or lysine in all eukaryotic and about half of archaeal sequences of RFC-A subunits. In subunits other than RFC-A, the residue at this position (for example, B-Thr120 in RFC-B) is either serine or threonine, residues that are commonly found at protein–DNA interfaces²⁹.

The strong conservation of these residues in the bacterial and archaeal clamp loaders suggests that the DNA-recognition model proposed here may be applicable to all clamp loader complexes. Indeed, comparison of the inactive form of the *E. coli* clamp loader complex^{11,30} with the RFC structure shows that the former can be described as a distorted and open form of the RFC spiral, which presumably closes into an RFC-like structure upon ATP and clamp binding (Fig. 6a).

The three distal helices ($\alpha 4$, $\alpha 5$, $\alpha 6$) of the RFC subunits correspond to the E, F and G helices of the canonical RecA fold³¹. In both RecA and the DnaB-type helicase from T7 bacteriophage, the two loops preceding the F and G helices (denoted L1 and L2, see Fig. 5) have been shown by mutagenesis to be involved in DNA recognition^{32–35}. In the Rho termination factor, a hexameric RNA helicase, two loops that are implicated in binding to single-stranded RNA within the central region of the hexamer correspond to the L1 and L2 loops of RecA³⁶. RecA forms spiral filaments with a significantly larger pitch than the RFC spiral (~ 14 Å per 60° turn, compared with ~ 5.6 Å in RFC)³¹. DnaB-type helicases and Rho form closed rings, and also spiral assemblies (but with open ATP-binding sites)^{36–38}. Despite these differences in gross subunit organization, it appears that the DNA-recognition mode that we propose for RFC is a variation on a theme that is common to RecA-type ATPases that interact with DNA or RNA.

A mechanism for primer–template recognition

The efficiency of lagging strand synthesis is increased by the ability of clamp loaders to position sliding clamps such that they are oriented appropriately with respect to the 3' end of the primer strand¹⁸. Footprinting experiments using human RFC have demonstrated that RFC recognizes DNA structures with recessed 3' ends specifically, in a sequence-independent manner, and interacts with both double- and single-stranded DNA across the primer–template junction³⁹. Our model, which takes primed DNA through the hole in PCNA and into the RFC spiral, fits well with the footprinting data and provides a simple mechanism for distinguishing between

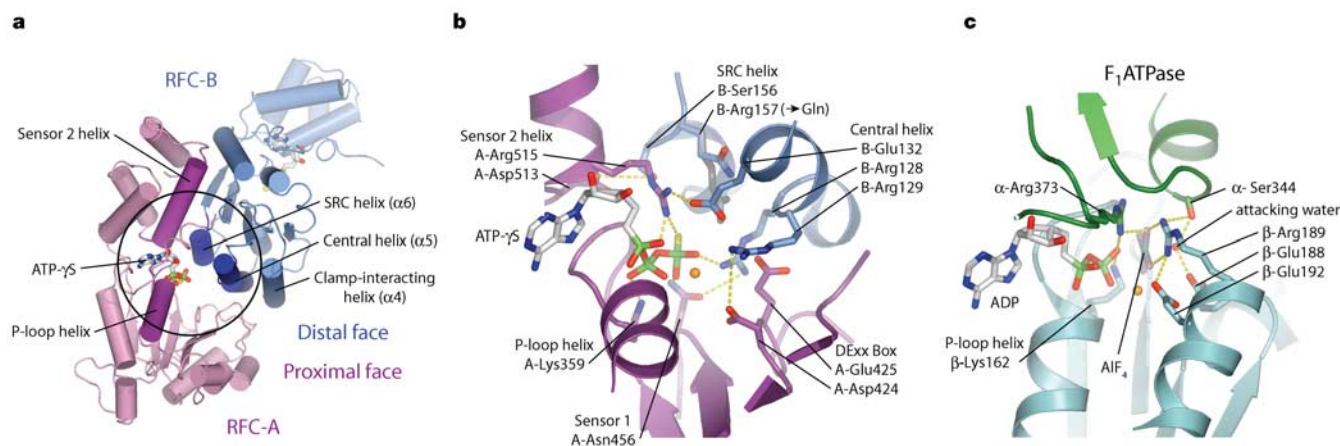
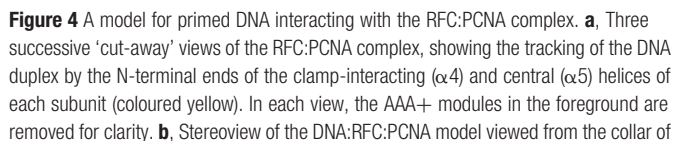


Figure 3 Nucleotide is tightly bound between proximal and distal faces of the RFC-A and RFC-B AAA+ modules. **a**, The RFC-A:RFC-B interface. **b**, Key ATP-interacting residues at the active site of RFC-A. Located near the terminal phosphate group of ATP- γ S is the arginine finger residue of the SRC motif of RFC-B (B-Arg157). This position has been

mutated to glutamine in this structure, and one possible conformation for an arginine side chain at this position is shown in grey. **c**, A view of the β_{DP} site in F_1 ATPase²⁴ (accession code 1H8E).

If we position RFC-A alongside the minor groove of the ideal DNA duplex (as shown in Fig. 4a), then the strand that runs in the 5' to 3' direction along the RFC spiral (the primer strand) ends up running into the interior wall of the RFC complex, near RFC-E. In contrast, the strand that runs in a 3' to 5' direction (the template strand) ends up facing the gap between RFC-E and RFC-A, where there appears to be sufficient room for a 5' extension to snake out of

To illustrate this directional specificity in the recognition of primed DNA, we use the structure of an actual primer–template junction, as seen in the crystal structure of HIV reverse transcriptase trapped to primed DNA⁴⁰ (Fig. 6b). The ~8 base pairs of the double helix preceding the 3'-primer terminus, which is bound to the active site of reverse transcriptase, are A-form in this structure and ~13 base pairs of the duplex that are further upstream are in a conformation resembling B-form. As expected, the junction between the B-form and the A-form DNA involves a bend (~40°) in the direction of the helix axis. If we align the A-form portion of the DNA along the axis of the RFC spiral, the preceding B-form



RFC, with the collar removed. A potential exit path for the 5' end of the template strand is indicated by green spheres. **c**, Schematic representation of the DNA:RFC:PCNA model. Alignment of RFC-A with the minor groove of the double helix positions the 5' terminus of the template strand near the opening between RFC-E and RFC-A.

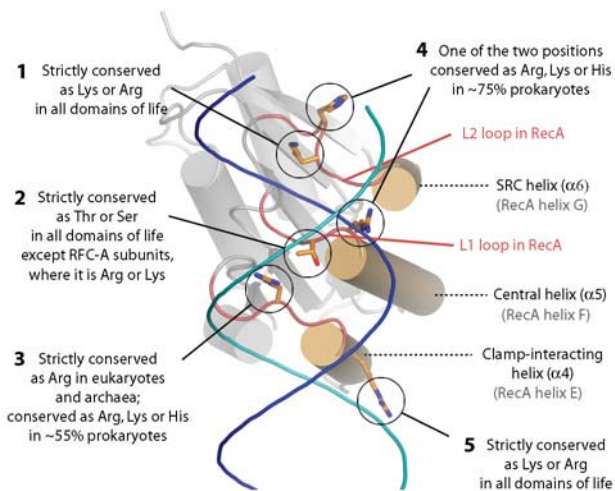


Figure 5 Conserved residues in domain I of clamp loader subunits at the proposed DNA-interacting surface. The loops preceding the central ($\alpha 5$) and SRC ($\alpha 6$) helices are equivalent to the L1 and L2 loops of RecA³¹, which have been implicated in DNA binding through extensive mutagenesis of RecA and DnaB-type helicases^{32–35}. Conserved positions are indicated on the basis of sequence analysis of 24 eukaryotic (16 small (for example, RFC-B) and 8 large (for example, RFC-A)), 35 archaeal (18 small and 17 large) and 63 prokaryotic (homologous to the *E. coli* γ -subunit) clamp loader subunits. Residue numberings for *S. cerevisiae* RFC-B (RFC4) and the *E. coli* γ -subunit are as follows: (1) B-Lys149, γ -Lys161; (2) B-Thr120, γ -Ser132; (3) B-Arg84, γ -Arg98/ γ -Lys100; (4) γ -Arg133 and γ -Gln160; (5) B-Arg90, γ -Arg105.

segment fits comfortably within the PCNA ring, which is large enough to accommodate the bend in the DNA. The single-stranded portion of template extends out of the gap in the RFC spiral, near domain IV of RFC-A. The conservation of positive charge within a surface groove of RFC and the bacterial clamp loaders suggests a possible exit path for the template strand (Fig. 4b).

Conclusions

By introducing mutations into the active sites of the RFC complex, we have captured the clamp loader loaded with an ATP analogue and bound to the closed PCNA sliding clamp. The arrangement of the ATPase domains within the complex suggests that these domains spiral around double-helical primed DNA. The RFC-A subunit is positioned near the mouth of the PCNA ring, enabling it to bind readily to DNA emerging from the clamp. If we assume that this interaction involves the minor groove of the double helix, then the length of the 'readout' of DNA by the RFC spiral (~ 11 base pairs) is such that the 3' end of the double helix must terminate within the RFC complex. This condition provides a simple mechanism for the specific recognition of primer-template junctions by the RFC complex.

The geometry of the double helix appears to match the geometry of an RFC spiral that generates optimally configured active sites, as seen at the RFC-A:RFC-B interface. Just as the central coiled-coil shaft of F_1 ATPase (the γ -subunit) is coupled to ATP hydrolysis, DNA may serve to stabilize the ATPase domains of RFC in configurations that are optimal for catalysis. Comparison with the structures of RecA, the replicative helicase from T7 bacteriophage and the Rho termination factor suggests that an ancient mechanism for DNA recognition by RecA-type ATPases has been tailored by the clamp loaders for the specific loading of sliding clamps onto primer-template junctions. Our structure of the RFC:PCNA complex is a first view of an AAA+ ATPase assembly engaging one of its substrates, and we look forward to being able to compare the clamp loader mechanism with that of other members of this extensive superfamily.

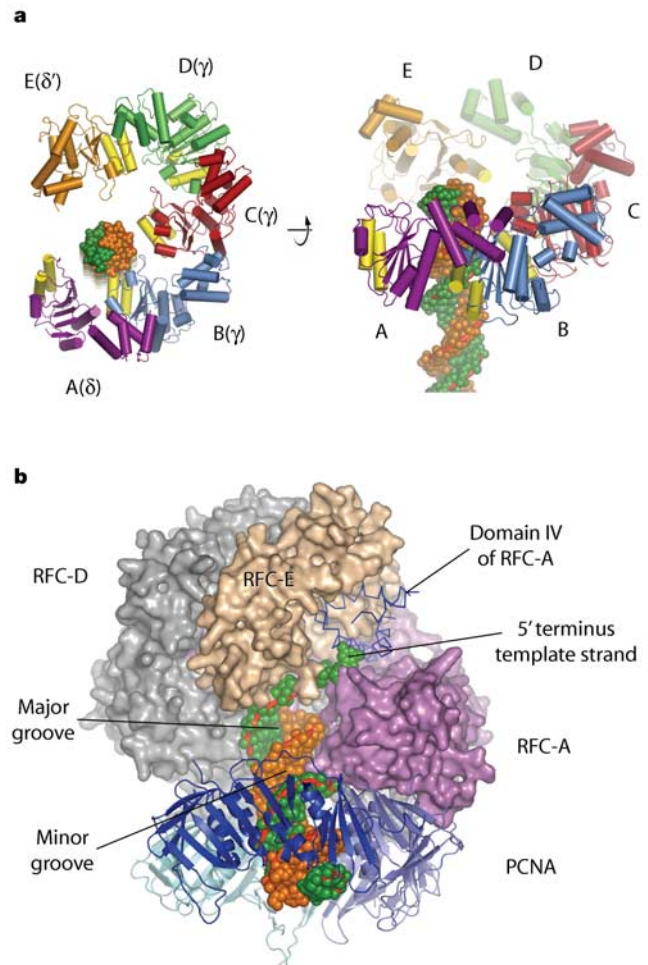


Figure 6 Inactive and active clamp loader complexes. **a**, The nucleotide-free, inactive state of the *E. coli* clamp loader¹¹ appears to be organized into a proto-spiral. Two views of the clamp loader are shown, oriented with respect to ideal B-form DNA by superimposing the B subunit of the *E. coli* clamp loader onto the B subunit of the RFC:PCNA:DNA model. The five *E. coli* subunits are coloured according to their counterparts in the RFC complex, and the collar has been removed for clarity. **b**, The RFC:PCNA complex modelled with primer-template DNA from the crystal structure of HIV reverse transcriptase⁴⁰ (accession code 1RTD). The primer and template strands are in orange and green, respectively. The molecular surface of the RFC complex is shown, except for domain IV of RFC-A.

Note added in proof: A three-dimensional reconstruction of the RFC:PCNA complex from the archaeobacterium *Pyrococcus furiosus* has been obtained using cryo-electron microscopy⁴⁹. □

Methods

Protein preparation

S. cerevisiae RFC and PCNA were expressed and purified as described in Supplementary Information. A truncated construct of RFC-A (residues 295–785) and full-length but mutant forms of the other subunits (RFC-B (R157Q), RFC-C (R160Q), RFC-D (R183Q) and RFC-E (R184Q); for nomenclature, see Fig. 2) were combined to form the RFC complex and purified as for wild-type RFC. The RFC complex was mixed with a ~ 2 molar excess of trimeric PCNA, 10 mM $MgCl_2$ and 0.5 mM ATP- γ S, concentrated, and passed over a Superdex-200 (10/30) column (Amersham Biosciences) that was equilibrated in 200 mM NaCl, 20 mM HEPES, pH 7.6, 0.5 mM ATP- γ S, 10 mM $MgCl_2$ and 5 mM β -mercaptoethanol (β -ME). Fractions of RFC:PCNA were then concentrated and frozen in small aliquots for crystallization.

Crystal growth and handling

Crystals were grown at 20 °C by the hanging drop method, using equal volumes of protein and crystallization solutions over a 500 μ l reservoir. Before mixing RFC:PCNA (30–50 mg ml^{-1}) with the reservoir solution (300 mM NaCl, 14.4–15.2% w/v polyethylene glycol 3350 (PEG 3350), 100 mM 2-(N-cyclohexylamino)ethanesulphonic acid (CHES), pH 9.0, 5 mM β -ME), final concentrations of 10 mM ATP- γ S, 10 mM $MgCl_2$, 4 mM

DL-dithiothreitol (DTT), 2 mM Tris (2-carboxyethyl)phosphine hydrochloride (TCEP) were added to the protein solution. Crystals were propagated by streak-seeding at the time of set-up, and were visible within several hours. The diffraction limit for these crystals was no better than ~ 5 Å, but was significantly improved by controlled crystal dehydration¹⁹. Two to four days after set-up, crystals measuring approximately $50 \times 100 \times 200$ μm were harvested into a sitting drop reservoir containing 50 μl mother liquor plus 1 mM ATP-γS and 10 mM MgCl₂, and the PEG 3350 concentration was steadily increased by slowly replacing the solution over 2 hours with an identical mix containing 33% PEG 3350. Crystals were then plunged into freshly thawed propane slush and stored in liquid nitrogen.

Data collection and structure determination

Diffraction data were measured at the Advanced Light Source, Lawrence Berkeley National Laboratory, on beamline 8.2.2 and processed with HKL2000 Denzo/Scalepack⁴¹. RFC:PCNA crystallizes in the spacegroup $P2_12_12_1$, with one complex in the asymmetric unit. Dehydration shrinks the unit cell from $a \sim 102.6$ Å, $b \sim 116.8$ Å, $c \sim 290.1$ Å to $a = 104.2$ Å, $b = 110.4$ Å, $c = 268.2$ Å. Phases were calculated from a two-wavelength anomalous dispersion (MAD) experiment at the selenium inflection and high remote energies⁴² using an RFC:PCNA crystal in which only the PCNA component was labelled with selenomethionine, and 17 selenium atoms were located using SOLVE⁴³. Electron density maps were calculated using MAD phases to 3.8 Å resolution, and greatly improved with solvent flattening and phase extension to 3.2 Å using SHARP⁴⁴ and DM⁴⁵. A further, isomorphous data set was collected to 2.85 Å resolution on a fresh crystal and combined with the MAD phases.

Model building at 2.85 Å resolution was performed with the program O⁴⁶, and refinement was carried out using CNS⁴⁷. The R_{work} and R_{free} values are 0.250 and 0.306, respectively. The PCNA subunits contacting RFC and the AAA+ modules of RFC-A, RFC-B and RFC-C are the best defined (average B -values ranging from 56 to 79 Å²), whereas the RFC-D and RFC-E subunits and domain IV of RFC-A exhibit significantly more disorder (average B -values ranging from 109 to 131 Å²). In particular, the unique domain IV of RFC-A is built as a polyalanine chain for much of its length. Molecular transformations were calculated using TOP⁴⁸ and FIT (<http://bioinfo1.mbfys.lu.se/~guoguang>).

Received 27 February; accepted 20 April 2004; doi:10.1038/nature02585.

- Huang, C. C., Hearst, J. E. & Alberts, B. M. Two types of replication proteins increase the rate at which T4 DNA polymerase traverses the helical regions in a single-stranded DNA template. *J. Biol. Chem.* **256**, 4087–4094 (1981).
- Prelisch, G., Kostura, M., Marshak, D. R., Mathews, M. B. & Stillman, B. The cell-cycle regulated proliferating cell nuclear antigen is required for SV40 DNA replication *in vitro*. *Nature* **326**, 471–475 (1987).
- Stukenberg, P. T., Studwell-Vaughan, P. S. & O'Donnell, M. Mechanism of the sliding β-clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.* **266**, 11328–11334 (1991).
- Kong, X. P., Onrust, R., O'Donnell, M. & Kuriyan, J. Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* **69**, 425–437 (1992).
- Krishna, T. S., Kong, X. P., Garry, S., Burgers, P. M. & Kuriyan, J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **79**, 1233–1243 (1994).
- Warbrick, E. The puzzle of PCNA's many partners. *Bioessays* **22**, 997–1006 (2000).
- Tsurimoto, T. & Stillman, B. Purification of a cellular replication factor, RF-C, that is required for coordinated synthesis of leading and lagging strands during simian virus 40 DNA replication *in vitro*. *Mol. Cell. Biol.* **9**, 609–619 (1989).
- Tsurimoto, T. & Stillman, B. Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. *Proc. Natl Acad. Sci. USA* **87**, 1023–1027 (1990).
- Lee, S. H., Kwong, A. D., Pan, Z. Q. & Hurwitz, J. Studies on the activator 1 protein complex, an accessory factor for proliferating cell nuclear antigen-dependent DNA polymerase δ. *J. Biol. Chem.* **266**, 594–602 (1991).
- Gomes, X. V. & Burgers, P. M. ATP utilization by yeast replication factor C. I. ATP-mediated interaction with DNA and with proliferating cell nuclear antigen. *J. Biol. Chem.* **276**, 34768–34775 (2001).
- Jeruzalmi, D., O'Donnell, M. & Kuriyan, J. Crystal structure of the processivity clamp loader γ complex of *E. coli* DNA polymerase III. *Cell* **106**, 429–441 (2001).
- Oyama, T., Ishino, Y., Cann, I. K., Ishino, S. & Morikawa, K. Atomic structure of the clamp loader small subunit from *Pyrococcus furiosus*. *Mol. Cell* **8**, 455–463 (2001).
- Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+ : A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
- Guenther, B., Onrust, R., Sali, A., O'Donnell, M. & Kuriyan, J. Crystal structure of the δ' subunit of the clamp-loader complex of *E. coli* DNA polymerase III. *Cell* **91**, 335–345 (1997).
- Yu, R. C., Hanson, P. I., Jahn, R. & Brunger, A. T. Structure of the ATP-dependent oligomerization domain of *N*-ethylmaleimide sensitive factor complexed with ATP. *Nature Struct. Biol.* **5**, 803–811 (1998).
- Lenzen, C. U., Steinmann, D., Whiteheart, S. W. & Weis, W. I. Crystal structure of the hexamerization domain of *N*-ethylmaleimide-sensitive fusion protein. *Cell* **94**, 525–536 (1998).
- Goedken, E. R. *et al.* Fluorescence measurements on the *E. coli* DNA polymerase clamp loader: implications for conformational changes during ATP and clamp binding. *J. Mol. Biol.* **336**, 1047–1059 (2004).
- Ason, B. *et al.* Mechanism of loading of the *Escherichia coli* DNA polymerase III β sliding clamp on DNA. Bona fide primer/templates preferentially trigger the γ complex to hydrolyze ATP and load the clamp. *J. Biol. Chem.* **278**, 10033–10040 (2003).
- Heras, B. *et al.* Dehydration converts DsbG crystal diffraction from low to high resolution. *Structure (Camb.)* **11**, 139–145 (2003).
- Shamoo, Y. & Steitz, T. A. Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* **99**, 155–166 (1999).
- Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M. & Kuriyan, J. Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell* **87**, 297–306 (1996).
- Lee, S. *et al.* The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell* **115**, 229–240 (2003).
- Niwa, H., Tsuchiya, D., Makyio, H., Yoshida, M. & Morikawa, K. Hexameric ring structure of the ATPase domain of the membrane-integrated metalloprotease FtsH from *Thermus thermophilus* HB8. *Structure (Camb.)* **10**, 1415–1423 (2002).
- Menz, R. I., Walker, J. E. & Leslie, A. G. Structure of bovine mitochondrial F₁-ATPase with nucleotide bound to all three catalytic sites: implications for the mechanism of rotary catalysis. *Cell* **106**, 331–341 (2001).
- Yang, W., Gao, Y. Q., Cui, Q., Ma, J. & Karplus, M. The missing link between thermodynamics and structure in F₁-ATPase. *Proc. Natl Acad. Sci. USA* **100**, 874–879 (2003).
- Waga, S. & Stillman, B. The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* **67**, 721–751 (1998).
- Saenger, W. *Principles of Nucleic Acid Structure* (Springer, New York, 1984).
- Xiong, Y. & Sundaralingam, M. Crystal structure of a DNA RNA hybrid duplex with a polypurine RNA r(gaagaagag) and a complementary polypyrimidine DNA d(CTCTTCTTC). *Nucleic Acids Res.* **28**, 2171–2176 (2000).
- Jones, S., van Heyningen, P., Berman, H. M. & Thornton, J. M. Protein–DNA interactions: A structural analysis. *J. Mol. Biol.* **287**, 877–896 (1999).
- Jeruzalmi, D. *et al.* Mechanism of processivity clamp opening by the δ subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. *Cell* **106**, 417–428 (2001).
- Story, R. M., Weber, I. T. & Steitz, T. A. The structure of the *E. coli* recA protein monomer and polymer. *Nature* **355**, 318–325 (1992).
- Hortnagel, K. *et al.* Saturation mutagenesis of the *E. coli* RecA loop L2 homologous DNA pairing region reveals residues essential for recombination and recombinational repair. *J. Mol. Biol.* **286**, 1097–1106 (1999).
- Mirshad, J. K. & Kowalczykowski, S. C. Biochemical characterization of a mutant RecA protein altered in DNA-binding loop 1. *Biochemistry* **42**, 5945–5954 (2003).
- Notarnicola, S. M., Park, K., Griffith, J. D. & Richardson, C. C. A domain of the gene 4 helicase/primase of bacteriophage T7 required for the formation of an active hexamer. *J. Biol. Chem.* **270**, 20215–20224 (1995).
- Washington, M. T., Rosenberg, A. H., Griffin, K., Studier, F. W. & Patel, S. S. Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding. *J. Biol. Chem.* **271**, 26825–26834 (1996).
- Skordalakes, E. & Berger, J. M. Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* **114**, 135–146 (2003).
- Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C. & Ellenberger, T. Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell* **99**, 167–177 (1999).
- Singleton, M. R., Sawaya, M. R., Ellenberger, T. & Wigley, D. B. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**, 589–600 (2000).
- Tsurimoto, T. & Stillman, B. Replication factors required for SV40 DNA replication *in vitro*. I. DNA structure-specific recognition of a primer–template junction by eukaryotic DNA polymerases and their accessory proteins. *J. Biol. Chem.* **266**, 1950–1960 (1991).
- Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669–1675 (1998).
- Otwiniowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Gonzalez, A. Optimizing data collection for structure determination. *Acta Crystallogr. D* **59**, 1935–1942 (2003).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- La Fortelle, E. D. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**, 476–494 (1997).
- Collaborative Computational Project Number 4, The CCP4 suite programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Kleywegt, G. J. & Jones, T. A. Efficient rebuilding of protein structures. *Acta Crystallogr. D* **50**, 829–832 (1996).
- Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Lu, G. TOP: A new method for protein structure comparisons and similarity searches. *J. Appl. Crystallogr.* **33**, 176–183 (2000).
- Miyata, T. *et al.* Direct view of the clamp-loading complex for processive DNA replication. *Nature Struct. Mol. Biol.* (in the press).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Berger, E. Goedken, M. Hingorani, D. Jeruzalmi, A. Johnson, S. Kazmirski, I. Nodelman, M. Podobnik, N. Yao and M. Young for scientific discussions, S. Chung and J. Finkelstein for technical assistance, K. Morikawa for sharing data prior to publication, L. Leighton for help with figure preparation, D. King for mass spectroscopy analysis, and members of the Kuriyan laboratory for support and helpful discussions. We appreciate the assistance of C. Ralston, G. McDermott and the scientific staff of the Advanced Light Source (LBNL, Berkeley) in data collection. This work was supported by grants from the National Institutes of Health (J.K. and M.O.D.) and a Ruth L. Kirschstein National Research Service Award through the National Institute of General Medical Sciences (G.D.B.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.K. (kuriyan@berkeley.edu). The coordinates and structure factors have been deposited in the Protein Data Bank (accession code 1SXJ).

A possible terrestrial analogue for haematite concretions on Mars

Marjorie A. Chan¹, Brenda Beittler¹, W. T. Parry¹, Jens Ormö² & Goro Komatsu³

¹Department of Geology and Geophysics, University of Utah, 135 S 1460 E, Salt Lake City, Utah 84112-0111, USA

²Centro de Astrobiología (INTA/CSIC), Instituto Nacional de Técnica Aeroespacial, Ctra de Torrejón a Ajalvir, km 4, 28850 Torrejón de Ardoz, Madrid, Spain

³International Research School of Planetary Sciences, Università d'Annunzio, Viale Pindaro 42, 65127 Pescara, Italy

Recent exploration has revealed extensive geological evidence for a water-rich past in the shallow subsurface of Mars. Images of *in situ* and loose accumulations of abundant, haematite-rich spherical balls from the Mars Exploration Rover 'Opportunity' landing site at Meridiani Planum^{1–3} bear a striking resemblance to diagenetic (post-depositional), haematite-cemented concretions found in the Jurassic Navajo Sandstone of southern Utah^{4,5}. Here we compare the spherical concretions imaged on Mars to these terrestrial concretions, and investigate the implications for analogous groundwater-related formation mechanisms. The morphology, character and distribution of Navajo haematite concretions allow us to infer host-rock properties and fluid processes necessary for similar features to develop on Mars. We conclude that the formation of such spherical haematite concretions requires the presence of a permeable host rock, groundwater flow and a chemical reaction front.

The existence of few, but conspicuous haematite (Fe₂O₃) deposits on Mars has been suggested from the Mars Global Surveyor Thermal Emission Spectrometer⁶ and are now confirmed by the instruments of the Opportunity Rover^{2,3}. In the 22-m-diameter Eagle Crater of the recent Opportunity landing site at Meridiani Planum, small (<0.5 cm), *in situ* haematite concretions are now reported in association with a high-albedo outcrop ledge (<1 m)^{1–3}. Mars haematite has previously been compared to terrestrial lake haematite occurrences^{6,7}, or spring fluids under ambient or hydrothermal conditions^{8,9}. The discovery of spherical haematite nodules^{2,3} and their accumulation patterns on Mars suggest an origin that may be similar to terrestrial examples of haematite concretions formed by groundwater flow^{10,11}.

Across much of southern Utah, the Jurassic Navajo Sandstone is one of the most porous and permeable formations^{4,12}, and easily transmits fluids as a reservoir and an aquifer. Cross-bedding is ubiquitous (Fig. 1a), and preserves the structure of wind-blown sand that accumulated on the lee (front) side of migrating desert dunes. The Navajo Sandstone host rock is a quartz arenite with weak quartz, carbonate and/or haematite cements, and relatively high porosity ~10–30%¹³. A remarkable abundance and variety of

haematite concretions (Fig. 1) are superbly exposed within the Navajo Sandstone.

Iron oxide as an indicator of fluid flow and occurrences of diagenetic haematite mineralization have been well documented in the Navajo Sandstone^{4,5,14–17} (Fig. 2). Fluid-flow history can be complicated with multiple generations of alteration, and with many colour combinations that indicate variable geochemical conditions over time. Near surface, meteoric waters and processes of weathering commonly distribute disseminated iron films that impart a pink to orange-red colour to the sandstone early in the depositional or burial history (Fig. 2a). The disseminated iron oxides are commonly mobilized and removed by reducing fluids (Fig. 2b), leaving the sandstone 'bleached' to pale, near-white colours^{4,13}. When these iron-rich fluids mix with oxidizing groundwater, haematite precipitates as a sandstone cement (Fig. 2c), typically in the form of spherical balls^{4,5}. These concretionary forms mostly consist of a minor mineral component (for example, iron oxides) that builds up a localized mass through diagenetic differentiation and precipitates as a pore-fill cement. Concretions differ in chemical composition from the matrix of their host rocks, and grains are tightly held by the strong and often pervasive cement that may comprise up to ~35% in whole-rock analysis (Fig. 2). In the Utah concretions, haematite is the major iron oxide, but other metastable iron oxides (for example, goethite-FeOOH) may also be present. As a consequence of the localized cement, concretions are typically more resistant to chemical and mechanical weathering than the host rock and 'weather out' to form an apron of loose concretions (Fig. 1a, b).

Field observations of numerous Navajo Sandstone concretion sites indicate abundant spherical forms, and different population sizes (millimetre to centimetre scale) (Figs 1d, 3). Many different and unusual shapes such as bulbous nodules, pipes and sheets are also common. Here we focus on the spherical forms because of their analogy and application to the Opportunity site. The spherical geometry is the minimum free-energy shape that forms in nature where the host rock is relatively homogeneous¹⁸ and there is no strong anisotropy such as joints, fractures, faults or other preferential fluid paths. Haematite concretions show a range of developmental stages. Weakly cemented (disseminated iron oxide) concretions with irregular outlines are commonly moderate to dark reddish-brown. Well-developed, smooth-surfaced concretions with more pervasive, homogeneous haematite cement are brownish-black (Fig. 1d) with larger grain sizes (>10 µm) and/or aggregates¹⁷. Concretion growth appears to be concentric, and typical haematite concretions display an outer cementation rind that may range from a thin 'egg shell' (<1 mm), to a thicker rind (>1 mm), to a series of nearly coalescing 'onion-skin' layers (Fig. 1d). Some concretions appear homogeneous or solid without any apparent rind (Fig. 1d).

Physical relationships of concretion size and spacing reflect host-rock characteristics as well as fluid chemistry, oxidation potential, pH, flow paths, and timing⁴. The presence of bacteria can enhance the rate of precipitation. In some geological settings, carbonate concretions typically form around some nucleus (a shell fragment,

Table 1 Comparison between characteristics of the Utah analogue and proposed characteristics of the system on Mars.

Utah		Mars
Host rock	Fine- to medium-grained (0.06–0.65 mm) quartz arenite	Fine-grained sedimentary ⁹
Fe source	Haematite grain coatings	Oxidized basalt (Fe ³⁺) ²³ , orthopyroxene (Fe ²⁺) ⁹ , other?
Fe mobilizing fluids	Reducing (± hydrocarbons)	CO ₂ (ref. 24), acidic brine ²⁵ , other?
Precipitating fluids	Oxidizing groundwater	Oxidizing groundwater
Concretion cements	Haematite (± goethite), carbonates (with no haematite)	Haematite ¹
Haematite characteristics		
<i>In situ</i> distribution	Internal within beds, along faults, fractures, joints	Internal ¹
Accumulations	Resistant surface concentrations, topographic lows	Surface layers, shallow depressions ¹
Geometry	Spherules + others based on host texture	Spherules ¹
Boundaries	Sharp to diffuse	Sharp ⁹ , diffuse?
Crystals	Fine (<10 µm) to coarse (>10 µm) pore-fill cement	Pure and crystalline ²

organic matter or a grain). Field observations and the lack of obvious macro nuclei in the Navajo concretions suggest a self-organized distribution for the spherical concretions, at some optimal nearest-neighbour spacing in zones along reaction fronts¹⁸. Small spherical concretion populations (millimetre size) are typically more closely spaced, and larger concretion populations (centimetre size) are more widely spaced (Figs 2c, 3). Individual concretion populations are fairly consistent with similar size and spacing in a given portion of a reaction front (Fig. 3). The concretion size, shape and spacing may also vary depending on the amount of fluids available and the growth time. Petrophysical characteristics can play a role in concretion size, so that adjacent sedimentary layers may have different concretion sizes or populations as a function of different porosity and permeability. Spherical concretions can be cemented across laminae, even where laminae are thinner than the diameter of the concretions. In these cases, anisotropy of individual lamina moderately affect the fluid flow,

and thus the spherical concretions may have a ridge-like feature around the periphery.

Mass-balance calculations indicate a range of fluid volumes to achieve the observed products. The amount of water needed is a function of pore and fluid chemistry, solubility and concretion-formation processes. Previous studies suggest that a range of 1,500–2,620 pore volumes^{15,16} are required to remove iron grain coatings ('bleaching') from regions of the Navajo Sandstone. Geochemical modelling suggests that a concretion containing 1 g of iron would require 100 kg of water, assuming a solubility of iron of about 10 mg kg^{-1} of water¹⁶. Extremes of pH (for example, acidic) will increase the solubility of iron and decrease the estimated pore volumes necessary. The terrestrial model of large pore volumes suggests that liquid groundwater in a permeable host rock and a history of fluid flow are required for haematite concretion formation. Groundwater flow might be driven by topography, and/or density gradients due to variations in temperature and/or salinity.

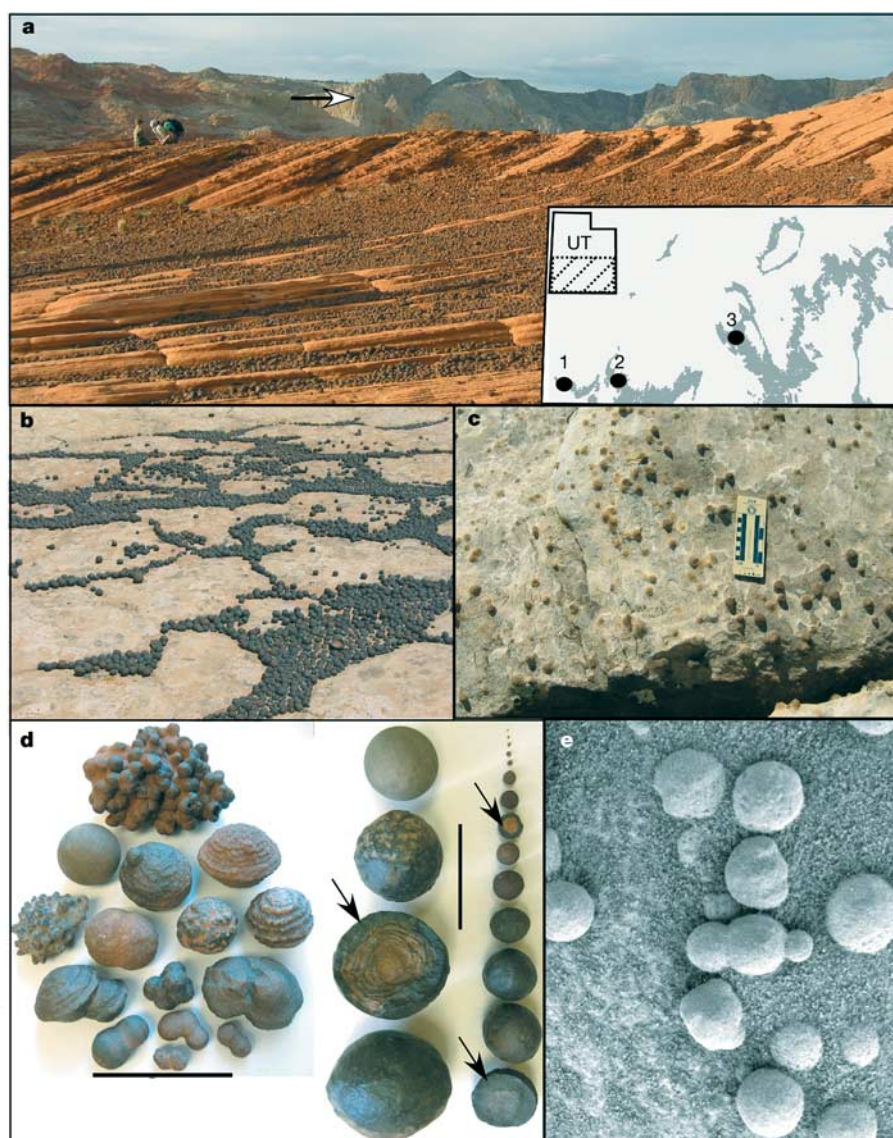


Figure 1 Utah haematite concretion examples (a–d) compared with Mars example (e). **a**, Iron mobility is reflected in sandstone colour and accumulated concretions along aeolian dune foresets. 'Bleached' sandstone in the distance (arrow), Locality 1. Inset shows state of Utah at upper left, with the region of southern Utah hashed, and shown in the larger inset. Shaded areas indicate exposures of Lower Jurassic outcrops. Localities: 1 = Snow Canyon 37.21° N, 113.65° W, 2 = Zion National Park 37.23° N, 112.89° W, and 3 = Grand Staircase Escalante National Monument 37.68° N, 111.37° W.

b, Weathered-out concretions of 2–3 cm diameter. **c**, *In situ* concretions. Scale card = 16.5 cm. **d**, A variety of concretions from different populations. Left, shapes. Right, sizes <1 mm–5 cm. Scale bars = 5 cm. Arrows, cross-sectional view. The lower right cross-section lacks an obvious rind. **b–d**, From Locality 3. **e**, Uncalibrated image of spherical nodules (≤ 0.5 cm) identified as haematite concretions¹ from the Opportunity site (1.98° S and 5.94° W)¹, presented only for Mars context. Image credit, NASA/JPL/Cornell.

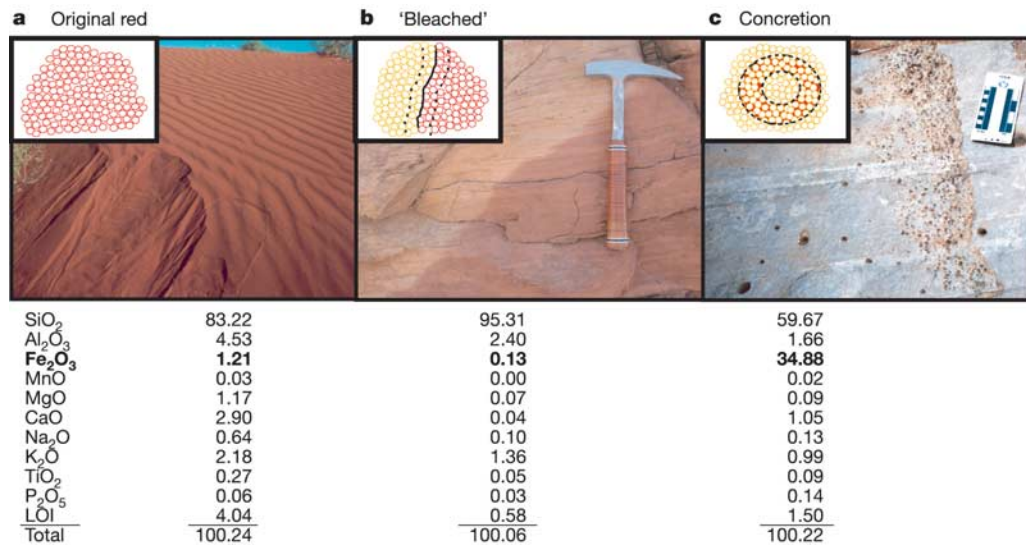


Figure 2 Model for Utah haematite concretion formation. Upper left insets show the formation at the grain scale⁵, with analogous outcrop examples of each stage, and representative inductively-coupled plasma mass spectrometry whole-rock analyses from Locality 3 are shown beneath. LOI = loss on ignition. **a**, Grains are coated by haematite films that provide an iron source for later concretions. Modern dune example from

Western Australia, courtesy of R. Ojakangas. (University of Minnesota, Duluth). **b**, After burial, iron is mobilized by reducing fluids that 'bleach' sandstone. Locality 3. **c**, Haematite concretions precipitate when iron-rich fluids meet with oxidizing groundwater. Scale card = 16.5 cm. Locality 2.

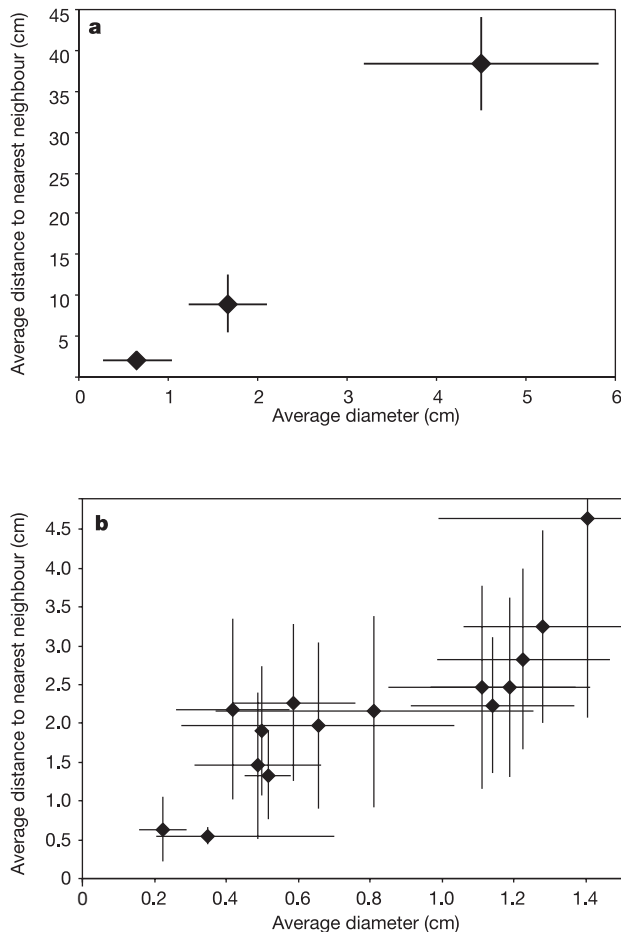


Figure 3 Navajo Sandstone statistics for haematite concretion size and spacing for select, small samples. The average diameter is plotted versus the average distance to the nearest neighbour for concretions. Error bars indicate one standard deviation. **a**, Linear trend between size and spacing for three populations of concretions at different portions of one reaction front ($n = 34$). Locality 2. **b**, Data from ten field sites. Small size populations from distinct reaction fronts ($n = 443$). Localities 1, 2, 3.

Concretions are known to form in rocks with diverse compositions including volcanic and sedimentary rocks, and the compositions of concretions can include iron oxides, carbonates, sulphides, silicates and many others. Iron concretions are not unique to Jurassic sandstones of Utah, and occur in a variety of geological settings, with documented examples from palaeosols to ocean and lake bottoms¹⁹. However, the spherical morphologies, distribution and haematite composition of the Navajo Sandstone examples bear a strong resemblance to the Mars concretions, and thus the well-documented characteristics make this an important terrestrial analogue. Similar well-studied goethite- and kaolinite-cemented concretions occur in Cretaceous sandstones in the Czech Bohemian Basin²⁰, with iron derived from igneous sources and distributed by heated water associated with intrusives. This shows that despite different iron sources and groundwater systems, in the same way that the Mars example might differ from the Utah concretions, the implications for subsurface fluid flow and porous host rock are similar.

The characteristics of Utah haematite concretions are compared to the reported observations from the Opportunity site in Table 1. The Utah analogue is similar to the Mars concretions in iron-rich fluids and oxidizing groundwater for precipitation of haematite mineralogy, *in situ* and accumulated distributions, and the small spherule geometry. There are clearly differences between the host-rock composition and concretion variability (more sizes, shapes and forms in the Utah examples). However, on the basis of the terrestrial model, we may expect other geometries of haematite concretions to be present on Mars as well. The interpretation of low-temperature ($\leq 100^\circ\text{C}$) during the formation of the Utah concretions is based on Navajo Sandstone burial estimates of $< 3\text{ km}$ (ref. 21) with a $50\text{--}70\text{ mW m}^{-2}$ constant basal heat flow²². The thermal gradient of the Navajo Sandstone is probably low because of the relatively high thermal conductivity (for example, $4\text{ W m}^{-1}\text{ K}^{-1}$)²². With an average of 60 mW m^{-2} and a thermal conductivity of $4\text{ W m}^{-1}\text{ K}^{-1}$, an estimated background thermal gradient is 15°C km^{-1} . For these regional thermal properties and burial history, and lacking other evidence of high-temperature mineral assemblages, we infer diagenetic temperatures. However, more work is needed to fully understand the role of temperature regimes in both the terrestrial and the

Mars concretion systems. Although the analogue is not a perfect match in every geologic parameter, the mere presence of these spherical haematite concretions implies significant pore volumes of moving subsurface fluids through porous rock.

Haematite is one of the few minerals found on Mars that can be linked directly to water-related processes. Utah haematite concretions are similar to the Mars concretions with spherical morphology, haematite composition, and loose, weathered accumulations. The abundance of quartz in the Utah example could pose challenges for finding spectral matches to the concretions. However, the similarities and differences of the Utah and Mars haematite concretions should stimulate further investigations. The potential role of biomediation in the precipitation of some terrestrial haematite concretions could also hold important clues in the search for life on Mars. Intriguing factors in the Mars system yet to be explored include: spectral comparisons, the source of the iron, the driving mechanism for fluid flow, and other chemical and physical parameters. This Utah analogue presents a model for a fascinating history of fluid flow in the haematite region of Mars. □

Received 18 February; accepted 26 April 2004; doi:10.1038/nature02600.

1. NASA Jet Propulsion Laboratory. *Mars Exploration Rover Mission* (<http://marsrovers.jpl.nasa.gov>) (February–March 2004).
2. Glotch, T. D. *et al.* Hematite at Meridiani Planum: Detailed spectroscopic observations and testable hypotheses. *Lunar Planet. Sci. Conf. XXXV* [online] Abstr. 2168; (<http://www.lpi.usra.edu/meetings/lpsc2004/>) (2004).
3. Squyres, S. W. *et al.* Initial results from the MER Athena Science investigation at Gusev Crater and Meridiani Planum. *Lunar Planet. Sci. Conf. XXXV* [online] Abstr. 2187; (<http://www.lpi.usra.edu/meetings/lpsc2004/>) (2004).
4. Chan, M. A., Parry, W. T. & Bowman, J. R. Diagenetic hematite and manganese oxides and fault-related fluid flow in Jurassic sandstones, southeastern Utah. *Am. Assoc. Petrol. Geol. Bull.* **84**, 1281–1310 (2000).
5. Chan, M. A. & Parry, W. T. Rainbow of rocks: mysteries of sandstone colors and concretions in Colorado Plateau Canyon Country. (Utah Geological Survey Public Information Service, Salt Lake City, Utah, 2002); (<http://www.ugss.state.ut.us/online/pdf/pi-77.pdf>).
6. Christensen, P. R., Morris, R. V., Lane, M. D., Bandfield, J. L. & Malin, M. C. Global mapping of Martian hematite mineral deposits: Remnants of water-driven processes on early Mars. *J. Geophys. Res.* **106**, 23873–23885 (2001).
7. Newsom, H. E. *et al.* Paleolakes and impact basins in southern Arabia Terra, including Meridiani Planum: Implications for the formation of hematite deposits on Mars. *J. Geophys. Res.* **108**, doi:10.1029/2002JE001993 (2003).
8. Catling, D. C. & Moore, J. M. The nature of coarse-grained crystalline hematite and its implications for the early environment of Mars. *Icarus* **165**, 277–300 (2003).
9. Hynek, B. M., Arvidson, R. E. & Phillips, R. J. Geological setting and origin of Terra Meridiani hematite deposit on Mars. *J. Geophys. Res.* **107**, doi:10.1029/2002JE001891 (2002).
10. Örmö, J. & Komatsu, G. Hydrocarbon related bleaching of strata and hematite deposition in red beds at Moab, Utah: A possible analogous process that formed bright layers and hematite deposits on Mars. *Lunar Planet. Sci. Conf. XXXIV* [online] Abstr. 1356; (<http://www.lpi.usra.edu/meetings/lpsc2003/>) (2003).
11. Beitley, B., Örmö, J., Komatsu, G., Chan, M. A. & Parry, W. T. Geomorphic and diagenetic analogs to hematite regions on Mars: Examples from Jurassic Sandstones of Southern Utah, USA. *Lunar Planet. Sci. Conf. XXXV* [online] Abstr. 1289; (<http://www.lpi.usra.edu/meetings/lpsc2004/>) (2004).
12. Antonellini, M. & Aydin, A. Effect of faulting on fluid flow in porous sandstones: geometry and spatial distribution. *Am. Assoc. Petrol. Geol. Bull.* **79**, 642–671 (1995).
13. Hood, J. W. & Patterson, D. J. *Bedrock Aquifers in the Northern San Rafael Swell Area, Utah, with Special Emphasis on the Navajo Sandstone*. Technical Publication 78 (Utah Department of National Research, 1984).
14. Chan, M. A., Parry, W. T., Petersen, E. U. & Hall, C. M. ⁴⁰Ar–³⁹Ar age and chemistry of manganese mineralization in the Moab to Lisbon fault systems, southeastern Utah. *Geology* **29**, 331–334 (2001).
15. Beitley, B., Parry, W. T. & Chan, M. A. Bleaching of Jurassic Navajo Sandstone on Colorado Plateau Laramide High: Evidence of exhumed hydrocarbon supergiants? *Geology* **31**, 1041–1044 (2003).
16. Parry, W. T., Chan, M. A. & Beitley, B. Chemical bleaching indicates fluid flow in sandstone deformation bands. *Am. Assoc. Petrol. Geol. Bull.* **88**, 175–191 (2004).
17. Beitley, B., Chan, M. A. & Parry, W. T. Field mapping and multispectral analysis of Jurassic Navajo Sandstone color and iron mineralization, Grand Staircase–Escalante National Monument, Utah. *Geol. Soc. Am. Abstr.* **34**, 277 (2002).
18. Ortoleva, P. T. *Geochemical Self-Organization* (Oxford Univ. Press, 1994).
19. Cornell, R. M. & Schwertmann, U. The iron oxides: Structures, properties, reactions, occurrences and uses. (VCH, New York, 1996).
20. Adamovic, J. in *Ironstones Pseudokarst Reports 2* (eds Adamovic, J. & Čilek, V.) 7–40 (Czech Speleological Society, Zlatý Kun, Prague, 2002).
21. Nuccio, V. F. & Condon, S. M. Burial and thermal history of the Paradox Basin, Utah and Colorado, and petroleum potential of the middle Pennsylvanian Paradox Formation. *US Geol. Surv. Bull.* **76**, O1–O41 (1996).
22. Morgan, P. & Gosnold, W. D. in *Geophysical Framework of the Continental United States* (eds Pakiser, L. C. & Mooney, W. D.) 493–522 (Geological Society of America Memoirs Vol. 172, Boulder, Colorado, 1989).
23. Minitti, M. E., Lane, M. D. & Bishop, J. L. A new hematite formation mechanism for Mars. *Lunar Planet. Sci. Conf. XXXV* [online] Abstr. 1999; (<http://www.lpi.usra.edu/meetings/lpsc2004/>) (2004).

24. Hoffman, N. White Mars: A new model for Mars' surface and atmosphere based on CO₂. *Icarus* **146**, 326–342 (2000).
25. Fernández-Remolar, D. *et al.* The Tinto River, an extreme acidic environment under control of iron, as an analog of the Terra Meridiani hematite site of Mars. *Planet. Space Sci.* **53**, 239–248 (2004).

Acknowledgements We thank the donors of the American Chemical Society Petroleum Research Fund, and the Bureau of Land Management–Grand Staircase Escalante National Monument for partial support of this research (to M.A.C. and W.T.P.). The work by J.O. was supported by the Spanish Ministry for Science and Technology and the Ramon y Cajal Program. The work by G.K. was supported by funding from the Italian Space Agency.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.A.C. (machan@mines.utah.edu).

Deterministic quantum teleportation with atoms

M. Riebe¹, H. Häffner¹, C. F. Roos¹, W. Hänsel¹, J. Benhelm¹, G. P. T. Lancaster¹, T. W. Körber¹, C. Becher¹, F. Schmidt-Kaler¹, D. F. V. James² & R. Blatt^{1,3}

¹Institut für Experimentalphysik, Universität Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria

²Theoretical Division T-4, Los Alamos National Laboratory, Los Alamos NM 87545, USA

³Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Technikerstraße 25, A-6020 Innsbruck, Austria

Teleportation of a quantum state encompasses the complete transfer of information from one particle to another. The complete specification of the quantum state of a system generally requires an infinite amount of information, even for simple two-level systems (qubits). Moreover, the principles of quantum mechanics dictate that any measurement on a system immediately alters its state, while yielding at most one bit of information. The transfer of a state from one system to another (by performing measurements on the first and operations on the second) might therefore appear impossible. However, it has been shown¹ that the entangling properties of quantum mechanics, in combination with classical communication, allow quantum-state teleportation to be performed. Teleportation using pairs of entangled photons has been demonstrated^{2–6}, but such techniques are probabilistic, requiring post-selection of measured photons. Here, we report deterministic quantum-state teleportation between a pair of trapped calcium ions. Following closely the original proposal¹, we create a highly entangled pair of ions and perform a complete Bell-state measurement involving one ion from this pair and a third source ion. State reconstruction conditioned on this measurement is then performed on the other half of the entangled pair. The measured fidelity is 75%, demonstrating unequivocally the quantum nature of the process.

Teleportation of a state from a source qubit to a target qubit requires three qubits: the sender's source qubit and an ancillary qubit that is maximally entangled with the receiver's target qubit, providing the strong quantum correlation. Once these states have been prepared, a quantum mechanical measurement is performed jointly on the source qubit and the ancilla qubit (specifically, a Bell-state measurement, which projects the two qubits onto a basis of maximally entangled states). In this process, the two qubits are projected onto one of four equally likely outcomes. At the same time, the non-local properties of quantum mechanics cause the target qubit to be projected onto one of four corresponding states,

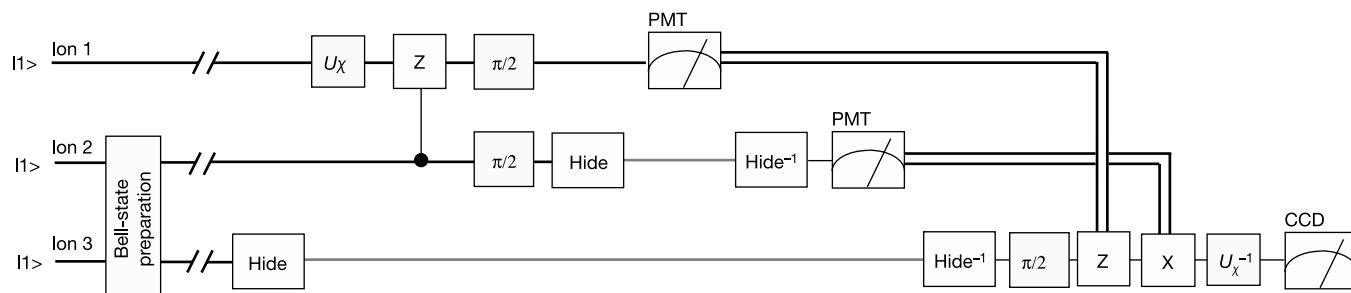


Figure 1 Teleportation from ion 1 to ion 3. A Bell state of ions 2 and 3 is prepared as a resource. The state to be teleported is encoded in ion 1 by the operation U_X . The Bell-state analyser consists of a controlled Z-gate followed by $\pi/2$ rotations and a state detection of ions 1 and 2. Note that this implementation uses a Bell basis rotated by $\pi/4$ with respect to the standard notation. Therefore a $\pi/2$ rotation on ion 3 is required before the reconstruction operations Z and X. The latter operations are realized by a π rotation

each related to the original state of the source qubit, even though no measurement was performed on this qubit. Knowledge of the result of the source-ancilla measurement allows one to choose a simple *a priori* operation to be carried out on the target qubit, resulting in reconstruction of the original quantum state. Because each of the four results on the source-ancilla measurement are equally likely, regardless of the nature of the teleported state, no information about the state is obtained (thus the no-cloning theorem⁷ is not violated); further, as classical communication of the measurement outcome is required to complete the state reconstruction, a state cannot be teleported faster than the speed of light. Teleportation does demonstrate a number of fascinating fundamental properties of quantum theory, in particular the non-local property of entangled states, which allows the projective measurement of the source-ancilla pair to create a definite pure state in the target qubit. Furthermore, teleportation has considerable implications for the nascent technology of quantum information processing⁸; besides being a compelling benchmark algorithm for a three-qubit quantum computer, teleportation is a possible primitive for large-scale devices⁹.

Experiments demonstrating qubit teleportation²⁻⁶, all of which used photons, were lacking the ability to perform a complete two-photon Bell-state measurement. Successful teleportation events were established by selecting the data after completion of the experiment, searching for the subset of experiments in which the outcome of the measurement and a preset reconstruction operation

were matched: that is, teleportation was performed post-selectively. One experiment¹⁰ demonstrated unconditional teleportation of continuous variables, in this case the quadrature amplitudes of a light field. Mention should be made of a liquid-state NMR experiment¹¹ in which an ensemble of molecules was used in a highly mixed state.

Our implementation of quantum teleportation uses entangled states of massive particles (see also ref. 12). For this, we store three $^{40}\text{Ca}^+$ ions in a linear Paul trap; for a detailed description of the experimental set-up, see ref. 13. The ions are arrayed in a linear crystal with an inter-ion distance of 5 μm . A qubit is encoded in a superposition of the $S_{1/2}$ ground state and the metastable $D_{5/2}$ state (lifetime $\tau \approx 1.16$ s) of a $^{40}\text{Ca}^+$ ion. Each qubit can be individually manipulated by a series of laser pulses on the $|1\rangle \equiv S_{1/2}$ ($m_J = -1/2$) to $|0\rangle \equiv D_{5/2}$ ($m_J = -1/2$) quadrupole transition near 729 nm employing narrow-band laser radiation tightly focused onto individual ions in the string. The qubits are initialized in $|1\rangle$ by optical pumping. The centre-of-mass vibrational mode ($\omega = 2\pi \times 1.2$ MHz) of the ion string is cooled to the ground state as required for controlled interaction between the ions¹³.

We teleport the quantum state of ion 1 (the source qubit) onto ion 3 (the target qubit) using the quantum circuit shown in Fig. 1. The sequence is formally equivalent to the one proposed by Bennett *et al.*¹, but adapted to the ion-based quantum processor. As a first step we prepare ion 2 (the ancilla) and 3 in the Bell state $|\psi_+\rangle_{23} = (|0\rangle_2|1\rangle_3 + |1\rangle_2|0\rangle_3)/\sqrt{2}$, the lifetime of which exceeds 100 ms

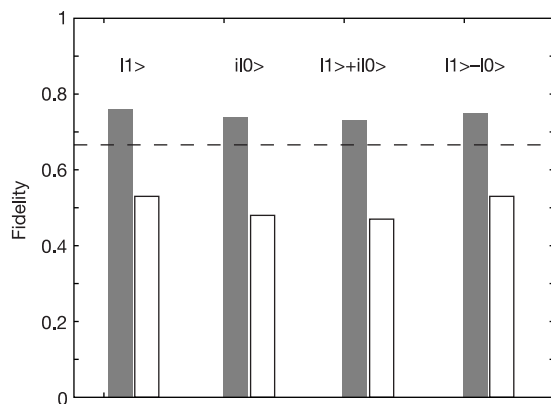


Figure 2 Result of the teleportation. The four test states are teleported with fidelities of 76%, 74%, 73% and 75%, respectively (grey bars). For each input state, 300 single teleportation experiments were performed. The error of each entry, estimated from quantum projection noise, is 2.5%. For comparison, white bars show the results if the reconstruction operations are omitted, yielding an average fidelity of 49.6%. The optimum 'teleportation' obtainable by purely classical means reaches a fidelity of 66.7% (dashed line).

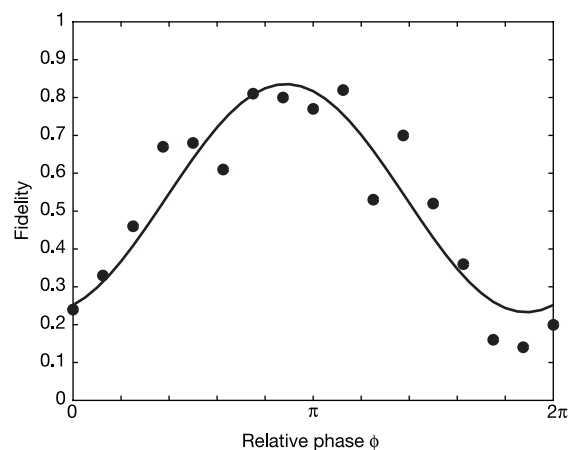


Figure 3 Fidelity of teleportation as a function of the relative phase ϕ of the reconstruction and analysing pulses (see Table 1). This is demonstrated for the test input state $|1\rangle$. Varying ϕ is used for optimizing the experimental fidelity thus accounting for a residual uncompensated phase drift during the entire process (which is the same for all input states).

(ref. 14). Then, at any time within this lifetime, the actual teleportation step—taking less than 2 ms—can be carried out: Ion 1 is prepared in an arbitrary input state with local rotations. In our experiments the teleported state $|\chi\rangle$ was one of a set of four non-orthogonal test states $|\chi^{(1)}\rangle = |1\rangle$, $|\chi^{(2)}\rangle = |0\rangle$, $|\chi^{(3)}\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $|\chi^{(4)}\rangle = (i|0\rangle + |1\rangle)/\sqrt{2}$. The Bell-state analysis is implemented by a controlled phase gate between ion 1 and 2 (ref. 11) followed by a $\pi/2$ pulse on each ion. Then the joint quantum state of ions 1 and 2 is measured by illuminating them with light at 397 nm for 250 μ s. Detection of fluorescence indicates the projection of the ion's quantum state into the $S_{1/2}$ state, (logical $|1\rangle$), whereas the absence of fluorescence indicates its projection into the $|0\rangle$ state with nearly 100% detection efficiency. Fluorescence is collected with a photomultiplier tube and the result is stored electronically. The measurement process must preserve the coherence of the target qubit, ion 3. Thus, the state of ion 3 is hidden by transferring it to a superposition of levels that are not affected by the detection light. In Ca^+ , an additional Zeeman level $|H\rangle \equiv D_{5/2}(m_J = -5/2)$ can be used for this purpose by moving any $S_{1/2}$ population into the $|H\rangle$ level. The same technique is used to sequentially read out ion 1 and ion 2 and discriminate between all four possible states $\{|0\rangle_1|0\rangle_2, |0\rangle_1|1\rangle_2, |1\rangle_1|0\rangle_2, |1\rangle_1|1\rangle_2\}$. Conditional on the measurement outcome, we apply the appropriate unitary qubit rotation, $-i\sigma_y$, $-i\sigma_z$, $i\sigma_x$ or 1 (with Pauli operators σ_k) to reconstruct the state in the target ion 3, obtaining $|\chi^{(\text{exp})}\rangle$. We emphasize that this conditional, deterministic step, in combination with the complete Bell-state analysis, is one of the crucial improvements with respect to all former experimental realizations of quantum teleportation with qubits. Furthermore, after the teleportation procedure the state $|\chi\rangle$ is always available and may be used for further experiments.

Table 1 Pulse sequence of the teleportation protocol.

	Action	Comment
1	Light at 397 nm	Doppler preparation
2	Light at 729 nm	Sideband cooling
3	Light at 397 nm	Optical pumping
Entangle		
4	$R_3^+(\pi/2, 3\pi/2)$	Entangle ion 3 with motional qubit
5	$R_2^C(\pi, 3\pi/2)$	Prepare ion 2 for entanglement
6	$R_2^C(\pi, \pi/2)$	Entangle ion 2 with ion 3
7	Wait for 1 μ s – 10,000 μ s	Standby for teleportation
8	$R_3^H(\pi, 0)$	Hide target ion
9	$R_1^H(\vartheta_\chi, \varphi_\chi)$	Prepare source ion 1 in state χ
Rotate into Bell basis		
10	$R_2^+(\pi, 3\pi/2)$	Get motional qubit from ion 2
11	$R_1^+(\pi/\sqrt{2}, \pi/2)$	Composite pulse for phasegate
12	$R_1^+(\pi, 0)$	Composite pulse for phasegate
13	$R_1^+(\pi/\sqrt{2}, \pi/2)$	Composite pulse for phasegate
14	$R_1^+(\pi, 0)$	Composite pulse for phasegate
15	$R_1^C(\pi, \pi/2)$	Spin echo on ion 1
16	$R_3^H(\pi, \pi)$	Unhide ion 3 for spin echo
17	$R_3^C(\pi, \pi/2)$	Spin echo on ion 3
18	$R_3^H(\pi, 0)$	Hide ion 3 again
19	$R_2^+(\pi, \pi/2)$	Write motional qubit back to ion 2
20	$R_1^C(\pi/2, 3\pi/2)$	Part of rotation into Bell basis
21	$R_2^C(\pi/2, \pi/2)$	Finalize rotation into Bell basis
Read out		
22	$R_2^H(\pi, 0)$	Hide ion 2
23	PM Detection for 250 μ s	Read out of ion 1 with photomultiplier
24	$R_1^H(\pi, 0)$	Hide ion 1
25	$R_2^H(\pi, \pi)$	Unhide ion 2
26	PM Detection for 250 μ s	Read out of ion 2 with photomultiplier
27	$R_2^H(\pi, 0)$	Hide ion 2
28	Wait 300 μ s	Let system rephase; part of spin echo
29	$R_3^H(\pi, \pi)$	Unhide ion 3
30	$R_3^C(\pi/2, 3\pi/2 + \phi)$	Change basis
Reconstruction		
31	$R_3^C(\pi, \phi)$	$i\sigma_x$
32	$R_3^C(\pi, \pi/2 + \phi)$	$-i\sigma_y$ } = $-i\sigma_z$ conditioned on PM detection 1
33	$R_3^C(\pi, \phi)$	$i\sigma_z$ conditioned on PM detection 2
34	$R_3^C(\vartheta_\chi, \varphi_\chi + \pi + \phi)$	Inverse of preparation of χ with offset ϕ
35	Light at 397 nm	Read out of ion 3 with camera

To obtain directly the fidelity of the teleportation, we perform on ion 3 the inverse of the unitary operation $U_\chi \equiv |\chi\rangle\langle 1| + |\bar{\chi}\rangle\langle 0|$, ($\langle\chi|\bar{\chi}\rangle = 0$) used to create the input state from state $|1\rangle$. The teleportation is successful if ion 3 is always found in $|1\rangle$. The teleportation fidelity, given by the overlap $F = \langle 1|U_\chi^{-1}\rho_{\text{exp}}U_\chi|1\rangle$, is plotted in Fig. 2 for all four test states. The obtained fidelities range from 73% to 76%. Teleportation based on a completely classical resource instead of a quantum-entangled resource yields¹⁵ a maximal possible average fidelity of 66.7% (dashed line in Fig. 2). Note, however, to rule out hidden variable theories, a fidelity in excess of 0.87 is required¹⁶. This level of fidelity could be reached by improving primarily the magnetic-field stability and the laser-frequency noise. For comparison, we also show data where the reconstruction pulses were not applied. Without the classical information communicated from the Bell measurement, the receiver's state is maximally mixed, that is there is no information available on the source state. Also, the measurement outcomes of ions 1 and 2 do not contain any information about the initial state $|\chi\rangle$. Indeed, we find each possible result with equal probability of 0.25 ± 0.036 , independent of the test input states. Note that only with both the receiver's qubit and the result of the Bell measurement, can the initial state be retrieved.

After the initial Bell-state preparation, only single qubit operations are performed on the target ion (see Fig. 1). Thus, although the scheme presented here takes place in a confining potential common to all three ions, it would, in principle, be possible to separate the target ion from the other two ions. This could be realized in segmented microtrap designs as developed previously¹⁷, which allow separating single ions from a crystal.

To illustrate further the quantum coherence of the teleportation process, we measure its fidelity as a function of the phase of the reconstructing and analysing pulses (see Fig. 3). The oscillation of the teleportation fidelity proves the phase relationship between source and target qubit and thus the quantum nature of the process.

To emphasize the role of the shared entangled pair as a resource, we store the Bell state for some time and then use it only later (after up to 10 ms) for teleportation. For waiting times up to 10 ms (exceeding the time we require for the teleportation by a factor of 5) we observe no decrease in the fidelity. Thus for future quantum-information processing networks, and with entangled states as a resource, quantum teleportation can be used for the distribution of quantum information between different nodes of the network. $\square$

Methods

Pulse sequence

To implement the teleportation we use pulses on carrier transitions $R_i^C(\vartheta, \varphi)$ and $R_i^H(\vartheta, \varphi)$ (no change of the motional state of the ion crystal) and, additionally, on the blue sideband $R_i^+(\vartheta, \varphi)$ (change of the motional state) on ion i . The index C denotes carrier transitions between the two logical eigenstates, and the index H labels the transition from the $S_{1/2}$ ($m_J = -1/2$) to the $D_{5/2}$ ($m_J = -5/2$) level in the Zeeman manifold. For the definitions of $R_i^{C,H,+}(\vartheta, \varphi)$ see refs 13 and 18. In contrast to Fig. 1, in the pulse sequence (Table 1) also the spin-echo pulses¹⁹ are included. These pulses make the algorithm more robust against detunings caused by slow laser and magnetic-field drifts. The timing of these generalized spin-echo pulses was found by optimizing the robustness of the sequence against detunings numerically and experimentally. The conditioned pulses 31,32,33 are applied only if less than six photon detection events were recorded during the respective detection time. For this purpose an electronic circuit counts the pulses of the photomultiplier tube and identifies the pulse sequence accordingly.

Received 17 March; accepted 16 April 2004; doi:10.1038/nature02570.

- Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and EPR channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
- Boschi, D., Branca, S., DeMartini, F., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
- Pan, J.-W., Daniell, M., Gasparoni, S., Weihs, G. & Zeilinger, A. Experimental demonstration of four-photon entanglement and high-fidelity teleportation. *Phys. Rev. Lett.* **86**, 4435–4438 (2001).
- Marcikic, I., de Riedmatten, H., Tittel, W., Zbinden, H. & Gisin, N. Long-distance teleportation of qubits at telecommunication wavelengths. *Nature* **421**, 509–513 (2003).
- Fattal, D., Diamanti, E., Inoue, K. & Yamamoto, Y. Quantum teleportation with a quantum dot single photon source. *Phys. Rev. Lett.* **92**, 037904 (2004).

7. Wootters, W. K. & Zurek, W. H. A single quantum cannot be cloned. *Nature* **299**, 802–803 (1982).
8. Nielsen, M. A. & Chuang, I. J. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
9. Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
10. Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
11. Nielsen, M. A., Knill, E. & Laflamme, R. Complete quantum teleportation using nuclear magnetic resonance. *Nature* **396**, 52–55 (1998).
12. Barrett, M. D. *et al.* Quantum teleportation with atomic qubits. *Nature* (this issue).
13. Schmidt-Kaler, F. *et al.* How to realize a universal quantum gate with trapped ions. *Appl. Phys. B* **77**, 789–796 (2003).
14. Roos, C. F. *et al.* Bell states of atoms with ultra long lifetimes and their tomographic state analysis. *Phys. Rev. Lett.* (in the press) Preprint at <http://arXiv.org/abs/physics/0307210> (2003).
15. Massar, S. & Popescu, S. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74**, 1259–1263 (1995).
16. Gisin, N. Nonlocality criteria for quantum teleportation. *Phys. Lett. A* **210**, 157–159 (1996).
17. Kielpinski, D., Monroe, C. & Wineland, D. J. Architecture for a large-scale ion-trap quantum computer. *Nature* **417**, 709–711 (2002).
18. Gulde, S. *et al.* Implementation of the Deutsch-Jozsa algorithm on an ion-trap quantum computer. *Nature* **412**, 48–50 (2003).
19. Hahn, E. L. Spin Echoes. *Phys. Rev.* **80**, 580–594 (1950).

Acknowledgements We thank H. Briegel and P. Zoller for a critical reading of the manuscript. We gratefully acknowledge support by the Austrian Science Fund (FWF), by the European Commission (QUEST, QUBITS and QGATES networks), by the Institut für Quanteninformation, and by the Los Alamos LDRD Program. This material is based upon work supported in part by the US Army Research Office. H.H. is funded by the Marie-Curie program of the European Union.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.B. (Rainer.Blatt@uibk.ac.at)

Deterministic quantum teleportation of atomic qubits

M. D. Barrett¹*, J. Chiaverini¹, T. Schaetz¹, J. Britton¹, W. M. Itano¹, J. D. Jost¹, E. Knill¹, C. Langer¹, D. Leibfried¹, R. Ozeri¹ & D. J. Wineland¹

¹Time and Frequency Division, NIST, Boulder, Colorado 80305, USA

²Mathematical and Computational Sciences Division, NIST, Boulder, Colorado 80305, USA

* Present address: Department of Physics, University of Otago, PO Box 56, Dunedin, New Zealand

Quantum teleportation¹ provides a means to transport quantum information efficiently from one location to another, without the physical transfer of the associated quantum-information carrier. This is achieved by using the non-local correlations of previously distributed, entangled quantum bits (qubits). Teleportation is expected to play an integral role in quantum communication² and quantum computation³. Previous experimental demonstrations have been implemented with optical systems that used both discrete and continuous variables^{4–9}, and with liquid-state nuclear magnetic resonance¹⁰. Here we report unconditional teleportation⁵ of massive particle qubits using atomic (⁹Be⁺) ions confined in a segmented ion trap, which aids individual qubit addressing. We achieve an average fidelity of 78 per cent, which exceeds the fidelity of any protocol that does not use entanglement¹¹. This demonstration is also important because it incorporates most of the techniques necessary for scalable quantum information processing in an ion-trap system^{12,13}.

Quantum teleportation¹ provides a means for transporting a quantum state between two separated parties, Alice and Bob, through the transmission of a relatively small amount of classical information. For the case of a two-state quantum system or ‘qubit’, only two bits of classical information are needed, which seems

surprising as precise specification of a general qubit state requires an infinite amount of classical information. Aside from the obvious differences in the various experimental demonstrations, the basic teleportation protocol is the same¹. Alice is in possession of a qubit (here labelled 2) that is in an unknown state $|\psi\rangle_2 \equiv \alpha|\uparrow\rangle_2 + \beta|\downarrow\rangle_2$, where $|\downarrow\rangle$ and $|\uparrow\rangle$ denote eigenstates of the qubit in the measurement basis. In addition, Alice and Bob each possess one qubit of a two-qubit entangled pair that we take to be a singlet $|S\rangle_{1,3} \equiv |\uparrow\rangle_1|\downarrow\rangle_3 - |\downarrow\rangle_1|\uparrow\rangle_3$ (where, for simplicity, we omit normalization factors). Therefore, Alice possesses qubits 1 and 2, while Bob holds qubit 3. Alice wishes to transmit the state of qubit 2 to Bob’s qubit using only classical communication. The initial joint state of all three qubits is

$$|\Phi\rangle = |S\rangle_{1,3} \otimes |\psi\rangle_2. \quad (1)$$

This state can be rewritten using an orthonormal basis of Bell states¹⁴ $|\Psi_k\rangle_{1,2}$ ($k = 1–4$) for the first two qubits and unitary transformations U_k acting on $|\psi\rangle_3 \equiv \alpha|\uparrow\rangle_3 + \beta|\downarrow\rangle_3$ so that $|\Phi\rangle = \sum_{k=1}^4 |\Psi_k\rangle_{1,2} (U_k |\psi\rangle_3)$. A measurement in the Bell-state basis $\{|\Psi_k\rangle\}$ by Alice then leaves Bob with one of the four possibilities $U_k |\psi\rangle_3$. Once Bob learns of Alice’s measurement outcome (through classical communication), he can recover the original unknown state by applying the appropriate unitary operator, U_k^{-1} , to his state $U_k |\psi\rangle_3$. We note that Alice’s Bell-state measurement can be accomplished by transforming from the basis $\{|\Psi_k\rangle_{1,2}\}$ into the measurement basis $\{|\uparrow\uparrow\rangle_{1,2}, |\uparrow\downarrow\rangle_{1,2}, |\downarrow\uparrow\rangle_{1,2}, |\downarrow\downarrow\rangle_{1,2}\}$ before the measurement.

Our implementation uses atomic qubits (⁹Be⁺ ions) that are confined in a linear radiofrequency Paul trap similar to that used in ref. 15. The control electrodes are segmented into eight sections as shown schematically in Fig. 1, providing a total of six trapping zones (centred on electrode segments 2 to 7). Potentials applied to these electrodes can be varied in time to separate ions and move them to different locations. The qubits are composed of the ground-state hyperfine levels $|\uparrow\rangle \equiv |F = 1, m = -1\rangle$ and $|\downarrow\rangle \equiv |F = 2, m = -2\rangle$, which are separated by $\omega_0 \equiv 2\pi \times 1.25$ GHz. These states are coupled through stimulated Raman transitions^{16–18} from two laser

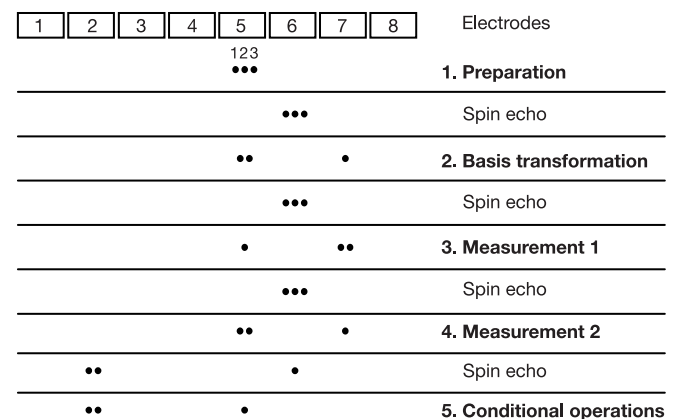


Figure 1 Schematic representation of the teleportation protocol. The ions are numbered left to right, as indicated at the top, and retain their order throughout. Positions, relative to the electrodes, are shown at each step in the protocol. The widths of the electrodes vary, with the width of the separation electrode (6) being the smallest at 100 μm . The spacing between ions in the same trap is about 3 μm , and laser-beam spot sizes (in traps 5 and 6) at the position of the ions are approximately 30 μm . In step 1 we prepare the outer ions in an entangled (singlet) state and the middle ion in an arbitrary state (equation (1)). Steps 2–4 constitute a measurement in a Bell-basis for ions 1 and 2 (Alice’s qubits), teleporting the state of ion 2 onto ion 3 (Bob’s qubit), up to unitary operations that depend on the measurement outcomes. In step 5 we invoke these conditional operations, recovering the initial state. Interspersed are spin-echo pulses applied in trap 6 that protect the state from de-phasing due to fluctuating magnetic fields but do not affect the teleportation protocol.

beams, which are used to implement the single-qubit rotations

$$R(\theta, \phi) = \cos(\theta/2)I + i\sin(\theta/2)\cos(\phi)\sigma_x + i\sin(\theta/2)\sin(\phi)\sigma_y, \quad (2)$$

where I is the identity operator, σ_x , σ_y and σ_z denote the Pauli spin matrices in the $\{|\uparrow\rangle, |\downarrow\rangle\}$ basis ($|\uparrow\rangle \equiv (1, 0)$, $|\downarrow\rangle \equiv (0, 1)$), θ is proportional to the duration of the Raman pulse, and ϕ is the relative phase between the Raman beams at the position of the ion. The Raman beams are also used to generate entanglement between two qubits by implementing the phase gate¹⁸

$$\begin{aligned} a|\uparrow\uparrow\rangle + b|\uparrow\downarrow\rangle + c|\downarrow\uparrow\rangle + d|\downarrow\downarrow\rangle \rightarrow & a|\uparrow\uparrow\rangle \\ & - ib|\uparrow\downarrow\rangle - ic|\downarrow\uparrow\rangle + d|\downarrow\downarrow\rangle. \end{aligned} \quad (3)$$

Our teleportation scheme is shown schematically in Fig. 1. We highlight the key elements of the protocol in bold lettering and also include the auxiliary ‘spin-echo’ pulses^{15,18} ($R(\pi, \phi_{SE})$) applied to ions in trap 6. These pulses are required in the experiment to prevent dephasing caused by variations in the ambient magnetic field on a timescale longer than the duration between the spin-echo pulses and, with an appropriate choice of ϕ_{SE} , can compensate phase accumulation due to the presence of a static magnetic-field gradient. As they do not fundamentally affect the teleportation, we omit their effects in the following discussion.

We first prepare the state $|S\rangle_{1,3} \otimes |\downarrow\rangle_2$ in two steps: starting from the state $|\downarrow\downarrow\downarrow\rangle_{1,2,3}$ we combine the gate in equation (3) applied to ions 1 and 3 with rotations to generate¹⁸ the state $(|\downarrow\downarrow\rangle_{1,3} - i|\uparrow\uparrow\rangle_{1,3}) \otimes |\downarrow\rangle_2$, followed by implementing individual ion rotations as discussed in ref. 19 to produce $|S\rangle_{1,3}$ from $|\downarrow\downarrow\rangle_{1,3} - i|\uparrow\uparrow\rangle_{1,3}$ (see methods section). For state $|S\rangle_{1,3} \otimes |\downarrow\rangle_2$, ions 1 and 3 are in the singlet, which is invariant under a global rotation. Therefore, a global rotation $R(\theta, \phi)_{1,2,3}$ to all three ions rotates the middle ion without affecting the singlet state of ions 1 and 3, and allows us to produce the state of equation (1) for any α and β with appropriate choices of θ and ϕ .

To teleport the state of ion 2 to ion 3, we start by implementing a Bell-state measurement on Alice’s qubits, ions 1 and 2. All three ions are transferred to trap 6 and then separated, with ions 1 and 2 going to trap 5 and ion 3 to trap 7. A phase gate (equation (3)) followed by a $\pi/2$ -pulse, $R(\pi/2, 0)$, is then applied to ions 1 and 2 in trap 5.

Our previous experiments¹⁵ showed a significant amount of motional-mode heating during the separation process, and the separation was achieved with only a 95% success rate. Aided by a smaller separation electrode in the current trap, we can separate the ions in the desired manner with no detectable failure rate. More importantly, the heating has been significantly reduced and we find that after the separation, the stretch mode of the two ions in trap 5 is in the ground state and the centre-of-mass mode has a mean quantum number of about 1. This enables us to implement the phase gate (equation (3)) between ions 1 and 2 with fidelity greater than 90% and without the need for sympathetic recoiling^{12,13,17}. Ideally (and in the absence of the spin-echo pulses) this leaves the ions in the state

$$\begin{aligned} & |\uparrow\uparrow\rangle_{1,2} \otimes R(\pi/2, -\pi/2)\sigma_x|\psi\rangle_3 + |\uparrow\downarrow\rangle_{1,2} \otimes R(\pi/2, -\pi/2)\sigma_y|\psi\rangle_3 \\ & + i|\downarrow\uparrow\rangle_{1,2} \otimes R(\pi/2, -\pi/2)I|\psi\rangle_3 - |\downarrow\downarrow\rangle_{1,2} \otimes R(\pi/2, -\pi/2)\sigma_z|\psi\rangle_3, \end{aligned} \quad (4)$$

where $|\psi\rangle_3 = \alpha|\uparrow\rangle_3 + \beta|\downarrow\rangle_3$.

To complete the Bell-state measurement, we need to detect the states of ions 1 and 2 individually. We recombine all three ions in trap 6 and separate them, with ion 1 being transferred to trap 5 and ions 2 and 3 transferred to trap 7. Detection on ion 1 is then achieved through state-dependent resonance fluorescence measurements¹⁹ ($|\downarrow\rangle$ strongly fluoresces whereas $|\uparrow\rangle$ does not), after which we optically pump the ion back to the state $|\downarrow\rangle_1$. All three ions are then recombined in trap 6 and separated again. For this separation, ions 1 and 2 are transferred to trap 5 and ion 3 is returned to trap 7. As the most recent spin-echo pulse applied in trap 6 transferred the

state of ion 1 to $|\uparrow\rangle_1$, a subsequent simultaneous detection of ions 1 and 2 effectively measures the state of ion 2 with error less than 1% due to the presence of ion 1.

To complete the teleportation, we apply unitary operations to ion 3 that depend on the measurement outcomes for ions 1 and 2. We first move ions 1 and 2 to trap 2 and ion 3 to trap 5, where unitary operations consisting of a $\pi/2$ -pulse, $R(\pi/2, \pi/2)$, followed by the operators σ_x , σ_y , I , σ_z for the measurement outcomes $|\uparrow\uparrow\rangle_{1,2}$, $|\uparrow\downarrow\rangle_{1,2}$, $|\downarrow\uparrow\rangle_{1,2}$, $|\downarrow\downarrow\rangle_{1,2}$ respectively, are applied. As noted above, the inclusion of the spin-echo pulses does not fundamentally change the teleportation protocol; however for $\phi_{SE} = \pi/2$, we must reorder the operations following the $\pi/2$ -pulse, $R(\pi/2, \pi/2)$, to I , σ_z , σ_x , σ_y respectively. A complete experiment is about 4 ms in duration, predominantly due to three elements: the cooling of all three axial modes to the ground state (1 ms), implementing the ion separations and movements (2 ms), and the three detection durations (0.6 ms). In the future, use of smaller trap electrodes to speed up ion-separation and gate operations, coupled with better detection, should considerably increase the speed of the teleportation process.

To demonstrate the full protocol we first teleport the basis states $|\uparrow\rangle_2$ and $|\downarrow\rangle_2$, and achieve a fidelity of about 80% ($78 \pm 3\%$ for $|\uparrow\rangle$ and $84 \pm 2\%$ for $|\downarrow\rangle$) for the data taken in the same run as that shown in Fig. 2). We also perform Ramsey experiments where the first $\pi/2$ pulse (having a variable phase ϕ) is applied to ion 2 (starting in the $|\downarrow\rangle_2$ state) and the second pulse (with a fixed phase) is applied to ion 3 after the teleportation is implemented. That is, $R(\pi/2, \phi)$ is applied to ion 2 and $R(\pi/2, \phi_{fixed})$ is applied to ion 3 after teleportation is completed. Ramsey fringes obtained in this way are shown in Fig. 2. We perform these experiments for $\phi_{fixed} = 0$ and $\pi/2$. From this data, we can extract the teleportation fidelities of the states $|\pm X\rangle$, $|\pm Y\rangle$, which are eigenstates of the operators σ_x and σ_y , respectively. From these fidelities and those for the states $|\uparrow\rangle$ and $|\downarrow\rangle$, we determine an average fidelity $\langle F \rangle = 78 \pm 2\%$ for the teleportation process. Furthermore, if we perform the teleportation without the conditional operations, the Ramsey fringes disappear and the teleportation fidelity drops to 1/2, equivalent to Bob making a random guess for the teleported state. There are three dominant mechanisms limiting the final fidelity; imperfect preparation of the initial state $|S\rangle_{1,3} \otimes |\downarrow\rangle_2$, imperfections in the second phase gate due to heating accrued in the separation process, and dephasing of the teleported state due to fluctuating magnetic fields. We have investigated these issues in independent experiments and find that each results in a loss of $8 \pm 2\%$ in the fidelity of the final state, consistent

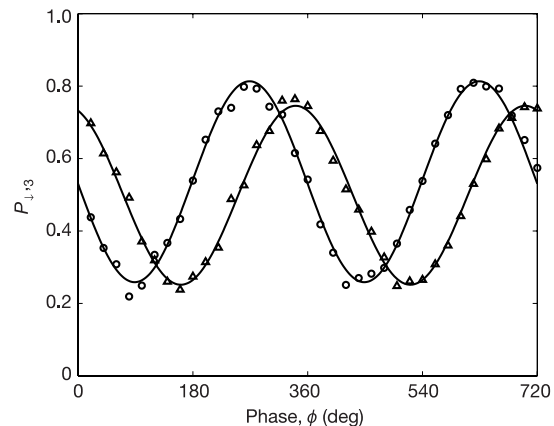


Figure 2 Ramsey fringes demonstrating the teleportation protocol. The two curves correspond to the second Ramsey pulse having $\phi_{fixed} = 0$ (circles) and $\phi_{fixed} = \pi/2$ (triangles) as discussed in the text. We plot the probability $P_{1,3}$ of observing ion 3 in the $|\downarrow\rangle_3$ state versus the phase of the first Ramsey pulse. Solid curves are best-fit sinusoidal functions to the data. The oscillations of the Ramsey fringes have an amplitude $|\rho_{11}|$ where $\rho_{11} = (\rho_{11})^*$ is the off-diagonal element of the density matrix of the teleported state. The fidelity of the teleported state is then given by $F = 1/2 + |\rho_{11}|$.

with the quoted result for teleportation. In the methods section, we discuss the effects of imperfections in the state preparation and teleportation process.

The average fidelity of $78 \pm 2\%$ achieved by our implemented quantum teleportation using atomic qubits, exceeds the value $2/3$ necessary to establish the presence of entanglement¹¹, and is accomplished on demand and without post-selection of the data. Although teleportation has been demonstrated in other systems, our demonstration incorporates the protocol into a simple experiment in such a way that it can be viewed as a subroutine of a quantum algorithm, here, a Ramsey experiment using two separated qubits. Furthermore, our demonstration incorporates most of the important features required for large-scale quantum information processing using trapped ions^{12,13}: We (a) reliably select qubits from a group and move them to separate trap zones while maintaining their entanglement, (b) manipulate and detect qubits without the need for strongly focused laser beams, and (c) perform quantum logic operations conditioned on ancilla measurement outcomes. Finally, we note that the University of Innsbruck has also implemented teleportation with the use of three Ca^+ ions in a linear Paul trap²⁰. $\square$

Methods

State preparation

Following the protocol outlined in ref. 18, three ions held in trap 5 are first laser-cooled, leaving all three axial modes in the ground state with 99% efficiency²¹. The internal states are then initialized to $|\downarrow\downarrow\downarrow\rangle$ by optical pumping. A motional phase-space 'displacement' pulse, inserted in a spin-echo sequence, is applied to the stretch mode of ions 1 and 3 by Raman beams that have a relative detuning of $\Delta\omega = \omega_s + \delta$, where ω_s is the frequency of the stretch mode and δ is a small detuning as described in ref. 18. The displacement pulse implements the transformation in equation (3) to ions 1 and 3. The spin-echo sequence acts on all three ions: a $\pi/2$ -pulse, a delay T_s in which the displacement pulse acts, a π -pulse, an equal time delay T_s , and a final $\pi/2$ -pulse. As the amplitude of motion for the middle ion is zero for the stretch mode, the entangling displacement pulse as described has no effect on this ion. Thus the spin-echo pulse sequence leaves the middle ion in the state $|\downarrow\rangle$, while the outer ions are affected as described in ref. 18. For an appropriate choice of detuning, δ , and pulse duration, $T = 2\pi/\delta$, we produce the state $(|\downarrow\downarrow\rangle_{1,3} - i|\uparrow\uparrow\rangle_{1,3}) \otimes |\downarrow\rangle_2$ (in the experiment, $T = 9.6 \mu\text{s}$).

To create the state $|\mathcal{S}\rangle_{1,3} \otimes |\downarrow\rangle_2$, a second spin-echo sequence is applied with the phase of the pulses shifted by $\pi/4$ with respect to the previous sequence. In addition, for the duration of the middle π -pulse, the axial confinement is changed. The resulting change in each ion's position gives rise to a relative phase of $-\pi/4$, 0 , and $+\pi/4$ for ions 1, 2 and 3, respectively¹⁹. For ion 2, the phase of the π -pulse is the same as the first and last $\pi/2$ -pulses and the sequence leaves the ion in the state $|\downarrow\rangle_2$. The outer two ions are initially in the state $|\downarrow\downarrow\rangle_{1,3} - i|\uparrow\uparrow\rangle_{1,3}$ and the first $\pi/2$ -pulse yields the state $|\uparrow\uparrow\rangle_{1,3} + |\downarrow\downarrow\rangle_{1,3}$. The π -pulse, with the $\pi/2$ phase difference on ions 1 and 3, results in the singlet state $|\uparrow\downarrow\rangle_{1,3} - |\downarrow\uparrow\rangle_{1,3}$. This state, being invariant under a global rotation, remains unchanged by the final $\pi/2$ -pulse. The three-ion state is then the desired state $|\mathcal{S}\rangle_{1,3} \otimes |\downarrow\rangle_2$. Auxiliary experiments establish a singlet fidelity $F_S = \text{Tr}_{2,3}(\langle \mathcal{S} | \rho_{\text{exp}} | \mathcal{S} \rangle) \approx 0.92(1)$ and a fidelity for the initial state of ion 2 of $\text{Tr}_{1,3}(\langle \downarrow | \rho_{\text{exp}} | \downarrow \rangle) \approx 0.95(1)$.

Imperfect state preparation and teleportation operations

As the operations in the experiment are imperfect, we must examine these effects to show that the observed teleportation fidelity that exceeds $2/3$ could not be caused by these imperfections. In particular, as the initial state of ions 1 and 3 is not a perfect singlet, when we rotate all three qubits to prepare the state $|\psi\rangle_2 = \alpha|\uparrow\rangle_2 + \beta|\downarrow\rangle_2$, we could potentially encode this information onto bit 3 even before the teleportation process is started. We address this issue as follows: We can verify that the gate in the state preparation does not entangle bit 2 with bits 1 and 3. Therefore the most general input state is given by

$$|\Psi\rangle_{\text{initial}} = (a_0|\downarrow\downarrow\rangle_{1,3} + a_1|\uparrow\uparrow\rangle_{1,3} + a_2|\uparrow\downarrow\rangle_{1,3} + a_3|\downarrow\uparrow\rangle_{1,3}) \otimes (b_0|\downarrow\rangle_2 + b_1|\uparrow\rangle_2), \quad (5)$$

where, ideally, $a_1 = -a_2 = 2^{-1/2}$ and $b_0 = 1$. Here we assume that the subsequent rotation to prepare $|\psi\rangle_2$ and the teleportation operations are perfect. However, in the experiment, after preparation of $|\Psi\rangle_{\text{initial}}$, we switch Raman beams from a geometry where the beams propagate at 90° to each other to a geometry where the beams co-propagate. Owing to fluctuating differences in the optical path lengths between these two sets of beams, there is a random phase difference from experiment to experiment between the corresponding Raman pulses. Over many experiments, some of the coherence terms between all three qubits average out due to this phase randomization, leading to a simplification of the calculated average fidelity with respect to the imperfections in state preparation.

The average fidelity $\langle F \rangle$ is evaluated from the expression

$$\langle F \rangle = (F(|\uparrow\rangle) + F(|\downarrow\rangle) + F(|X\rangle) + F(|-X\rangle) + F(|Y\rangle) + F(|-Y\rangle))/6, \quad (6)$$

where the arguments correspond to the initial state, $|\psi\rangle_2$. We find

$$\langle F \rangle = (2 - |b_0|^2)/3 + 2(2|b_0|^2 - 1)F_S/3. \quad (7)$$

When $F_S \leq 1/2$, $\langle F \rangle \leq 2/3$, and we conclude that the teleportation fidelity cannot exceed $2/3$ unless the teleportation has occurred through the entangled singlet state as

described in the text. Furthermore, imperfect state preparation of ion 2 contributes to an overall loss in fidelity. In a similar way, by simulation, we have also examined the effects of likely imperfections in the teleporting process and conclude that these imperfections only reduce the measured value of $\langle F \rangle$. Therefore, the experimentally measured value of $\langle F \rangle$ indicates that entanglement between bits 1 and 3 was required.

Received 17 March; accepted 4 May 2004; doi:10.1038/nature02608.

- Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- Briegleb, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: The role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
- Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
- Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
- Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
- Boschi, D., Branca, S., De Martini, F., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
- Kim, Y.-H., Kulik, S. P. & Shih, Y. Quantum teleportation of a polarization state with a complete Bell state measurement. *Phys. Rev. Lett.* **86**, 1370–1373 (2001).
- Bowen, W. P. *et al.* Experimental investigation of continuous-variable quantum teleportation. *Phys. Rev. A* **67**, 032302 (2003).
- Zhang, T. C., Goh, K. W., Chou, C. W., Lodahl, P. & Kimble, H. J. Quantum teleportation of light beams. *Phys. Rev. A* **67**, 033802 (2003).
- Nielsen, M. A., Knill, E. & Laflamme, R. Complete quantum teleportation using nuclear magnetic resonance. *Nature* **396**, 52–55 (1998).
- Massar, M. & Popescu, S. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74**, 1259–1263 (1995).
- Wineland, D. J. *et al.* Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Natl Inst. Stand. Technol.* **103**, 259–358 (2003).
- Kielipinski, D., Monroe, C. & Wineland, D. J. Architecture for a large-scale ion-trap quantum computer. *Nature* **417**, 709–711 (2002).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
- Rowe, M. A. *et al.* Transport of quantum states and separation of ions in a dual rf ion trap. *Quant. Inf. Comp.* **2**, 257–271 (2002).
- Wineland, D. J. *et al.* Quantum information processing with trapped ions. *Phil. Trans. R. Soc. Lond. A* **361**, 1349–1361 (2003).
- Barrett, M. D. *et al.* Sympathetic cooling of $^9\text{Be}^+$ and $^{24}\text{Mg}^+$ for quantum logic. *Phys. Rev. A* **68**, 042302 (2003).
- Leibfried, D. *et al.* Experimental demonstration of a robust, high-fidelity geometric two-ion-qubit phase gate. *Nature* **422**, 412–415 (2003).
- Rowe, M. A. *et al.* Experimental violation of a Bell's inequality with efficient detection. *Nature* **409**, 791–794 (2001).
- Riebe, M. *et al.* Deterministic quantum teleportation with atoms. *Nature* (this issue).
- King, B. E. *et al.* Cooling the collective motion of trapped ions to initialize a quantum register. *Phys. Rev. Lett.* **81**, 1525–1528 (1998).

Acknowledgements This work was supported by ARDA/NSA and NIST. We thank J. Bollinger and J. Martinis for helpful comments on the manuscript. T.S. acknowledges a Deutsche Forschungsgemeinschaft research grant. This paper is a contribution of the National Institute of Standards and Technology and is not subject to US copyright.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.W. (djw@boulder.nist.gov)

In situ observation of colloidal monolayer nucleation driven by an alternating electric field

Ke-Qin Zhang & Xiang Y. Liu

Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117542

The nucleation of crystalline materials is a hotly debated subject in the physical sciences¹. Despite the emergence of several theories in recent decades^{2–7}, much confusion still surrounds the dynamic processes of nucleation^{5–7}. This has been due in part to the limitations of existing experimental evidence. Charged

colloidal suspensions have been used as experimental model systems for the study of crystal nucleation and structural phase transitions^{8–12}, as their crystallization phase diagram is analogous to that of atomic and molecular systems, but they can be visualized using microscopy. Previously, three-dimensional imaging of colloidal nucleation dynamics was achieved using confocal microscopy¹³. However, the limited temporal resolution of the confocal microscope is of concern when trying to capture real-time colloidal crystal nucleation events. Moreover, as the thermodynamic driving force has remained undefined, data on key factors such as the critical nuclei size are at best semiquantitative¹³. Here we present real-time direct imaging and quantitative measurements of the pre- and post-nucleation processes of colloidal spheres, and the kinetics of nucleation driven by an alternating electric field, under well-defined thermodynamic driving forces. Our imaging approach could facilitate the observation of other rarely observed phenomena, such as defect and grain-boundary formation¹⁴ and the effects of foreign particles during crystallization¹⁵. Furthermore, it may prove useful in identifying optical and biological technologies based on colloids^{16–21}.

Nucleation is a dynamic process, in which small nuclei formed in a supersaturated liquid overcome a nucleation barrier. The growth

of these clusters depends on the competition between a decrease in bulk energy, which favours growth, and an increase in interfacial energy, which favours shrinkage. The smallest crystals are continually formed by fluctuations but then typically shrink away because of the interfacial energy. Growth becomes energetically favourable only when the crystallites reach a critical size. The nucleation rate and pre-nucleation distributions are determined by the thermodynamic driving force, $\Delta\mu/k_B T$, where $\Delta\mu$ is the difference in chemical potential between the growth units in the crystal and the liquid phase, k_B is Boltzmann's constant and T is the absolute temperature. Because this chemical potential difference is well defined in crystal growth, it can express the driving force in theoretical formulations of crystal growth. According to two-dimensional (2D) nucleation theories^{2–4}, the nucleation barrier (the cost of total Gibbs free energy ΔG_{crit} to form a critical circular nucleus with radius r_c in a supersaturated solution) is:

$$\Delta G_{\text{crit}} = \frac{\pi \Omega \gamma^2}{\Delta \mu} \quad (1)$$

$$r_c = \frac{\Omega \gamma}{\Delta \mu} \quad (2)$$

where γ is the crystal–liquid interfacial edge free energy (or line tension), and Ω is the area per structural unit. For a given

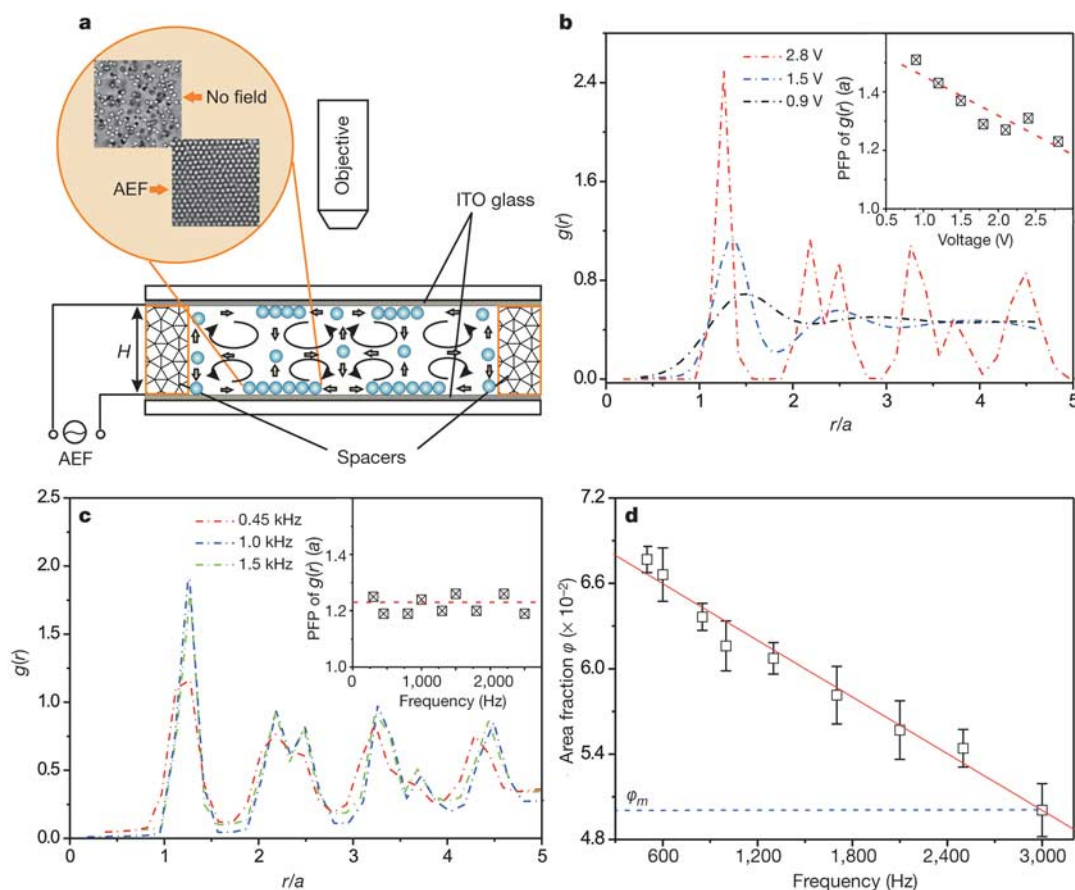


Figure 1 Method of assembling a 2D colloidal monolayer by controlling the alternating electric field (AEF). **a**, Sketch of the experimental set-up (thickness $H = 120 \pm 5 \mu\text{m}$). **b**, **c**, The structure of the 2D colloidal crystal is quantified by the pair correlation function (PCF), $g(r)$. The position of the first peak (PFP) of $g(r)$ corresponds to a fixed centre-to-centre distance between neighbouring particles. **b**, The effect on the colloidal crystal structure caused by varying the voltage from $V_{\text{pp}} = 0.9 \text{ V}$ to $V_{\text{pp}} = 2.8 \text{ V}$, while keeping the frequency fixed at $f = 600 \text{ Hz}$, is illustrated by plotting $g(r)$ against r/a at $V_{\text{pp}} = 2.8, 1.5, 0.9 \text{ V}$. The variation of the PFP of $g(r)$, which is independent of the shape of the peak with respect to V_{pp} , is shown in the inset by the slope of the plot of the first peak position

against the voltage. **c**, The absence of any effect on the colloidal crystal structure despite varying the frequency from $f = 450 \text{ Hz}$ to $f = 2,500 \text{ Hz}$, while keeping the field strength fixed at $V_{\text{pp}} = 2.4 \text{ V}$, is illustrated by plotting $g(r)$ against r/a at $f = 0.45, 1.0, 1.5 \text{ kHz}$. The constancy of the PFP of $g(r)$ with respect to the frequency is shown in the inset by the horizontal least-squares linear fit of the plot of the PFP against frequency. **d**, The supersaturation is examined by the fraction of the area of the electrode occupied by colloidal particles versus the frequency while keeping the voltage fixed at $V_{\text{pp}} = 2.4 \text{ V}$. Each data point is the average of five measurements; error bars show the standard deviation from the mean.

interparticle interaction, $\Delta\mu$ can be expressed²² in terms of the supersaturation σ :

$$\Delta\mu = k_B T \ln(1 + \sigma) \quad (3)$$

$$\sigma = \frac{\phi - \phi_m}{\phi_m} \quad (4)$$

where ϕ and ϕ_m represent respectively the actual concentration and the equilibrium concentration of the colloidal spheres in the coexistence phase.

The colloidal spheres are uniformly dispersed in water (see Methods) before the electric field is applied. Once the alternating electric field (AEF) is applied, the particles are transported to the surface of the electrode by the electric-field-induced fluid flow^{23–25}. Colloidal spheres self-assembled into a 2D hexagonal crystallite, formed at the bottom of the cell (Fig. 1a). The driving force for the assembly of 2D crystals is determined by the interaction among the colloidal spheres, and their local concentration. In order to obtain a well defined supersaturation based on equation (4) in describing the nucleation kinetics, we examine the effect of the peak-to-peak voltage V_{pp} and frequency f of the AEF (at a constant inter-electrode distance, H) on the area fraction, that is, the fraction of area covered by the colloidal particles on the surface of the electrode.

We notice that the position of the first peak (PFP) of the pair correlation function (PCF) $g(r)$ (ref. 26) is given by the density or average distance between particles and not their interactions. However, in the present context, the PCF seems to be dominated by particles in (small) crystals. In this case, changes in the interactions will change the crystal and thus the PCF. The variation of the PFP of $g(r)$ with increasing V_{pp} in Fig. 1b shows that V_{pp} (or the field strength E) changes the interaction between neighbouring particles. Because the PFP of $g(r)$ remains nearly the same (deviation <5%) at constant V_{pp} (compare Fig. 1c), the frequency f will not affect the interparticle interaction. A change in f will result in a change in the transport of the particles to the surface of the electrode and in the area/volume fraction of the colloidal particles (Fig. 1d). According to equation (3) and Fig. 1d, the driving force for colloidal crystallization can be changed by altering f . The dependence of the area fraction on frequency is given in Fig. 1d. For a given system, 2D colloidal crystals approach the melting stage at $f = 3.0$ kHz, when the area fraction of the colloidal particles corresponds to the

equilibrium point ϕ_m (see Methods). On the basis of Fig. 1d, the concentration of colloidal particles at the surface of the electrode can be controlled by altering f . By changing the frequency f of the AEF at a given V_{pp} , the thermodynamic driving force of equation (3) for the order–disorder phase transition of the colloidal monolayer can be controlled.

Figure 2a, b shows a typical process of nucleation and the evolution of some nuclei in two early time snapshots. Particles accumulated on the surface, and started to form nuclei. The crystal-like particles are highlighted by the blue spheres (see Methods). The nuclei in the red circles kept on growing, whereas the nuclei in the yellow circles shrank far more frequently than they grew (Fig. 2c, d). The statistical fluctuation of clusters at the pre- and post-nucleation stages was observed continuously by the light microscope.

Figure 3a shows a typical size distribution $Z_n(t)$ of 2D crystalline clusters at different stages of nucleation. As time t changes from $t_1 = 3.2$ s to $t_4 = 20.8$ s, the distribution $Z_n(t)$ approaches a steady state of nucleation (SSN) where $Z_n(t)$ is independent of t . The transition from t_1 to t_4 represents a transition from a transient state to a SSN. In the context of nucleation, the size distribution $C(n)$ for the equilibrium state can be estimated by the Boltzmann distribution^{3,27} based on $\Delta G(n) = 2\gamma(\pi\Omega n)^{1/2} - n\Delta\mu$, where γ , the line tension, was obtained from experiments based on equation (2). As predicted by theories^{3,27}, the experimental distribution $Z_n(t)$ of SSN is lower than the equilibrium state distribution $C(n)$, and will come closer to $C(n)$ as the cluster size n tends to 1. The critical nuclei are the crystalline clusters that overcome the nucleation barrier, and enter the stage of continuous growth at $n > n_c$. As the steady nucleation rate J is constant at a given supersaturation, the number of critical nuclei produced is constant with time. Owing to the fact that post-nucleation crystallites will grow continuously, the distribution of post-nucleation clusters $Z^*(t)$ for any $n > n_c$ should be constant. In other words, $Z^*(t)$ should be a horizontal line with respect to the n axis (Fig. 3a). The intercept between the pre-nucleation distribution $Z_n(t)$ and the post-nucleation distribution $Z^*(t)$ gives rise to a critical cluster size $n_c(t)$ (Fig. 3a inset) at a given moment t .

Figure 3b shows that $n_c(t)$ increases with time and reaches a constant value n_c at $t > \tau$. Here τ is the induction time for the transition from the non-steady state to the steady state, and the constant value n_c is the critical size of a 2D nucleus at the SSN at a given supersaturation. Thus we can determine the critical size n_c of SSN, the critical radius r_c of a nucleus, the line tension γ and nucleation barrier ΔG_{crit} at different supersaturations (see Supplementary Table).

The nucleation kinetics are characterized by a nucleation rate, which is defined as the number of ‘mature’ nuclei ($r \geq r_c$) created per unit time per unit area. Current nucleation theories²⁷ mainly focus on the SSN. As illustrated in Fig. 3a and b, our approach enables us to examine this well defined state at precisely controlled supersaturations. According to the definition, the total number N of ‘mature’ nuclei ($n > n_c$) at a given area is plotted versus time t in Fig. 3c, which gives rise to a linear relation between N and t . The slope of the N versus t line corresponds to the steady nucleation rate J . According to the 2D nucleation models²⁸, the nucleation rate of the SSN can be expressed as a function of supersaturation as:

$$J = C \left\{ \frac{2D_s n_1^2}{\pi} \left[\frac{\Omega \ln(1 + \sigma)}{a} \right]^{1/2} \exp \left(- \frac{\pi \Omega \gamma^2}{(k_B T)^2 \ln(1 + \sigma)} \right) \right\} \quad (5)$$

Here D_s is the diffusion coefficient, n_1 is the number of single particles (monomers), C is the kinetic coefficient and a is the diameter of the colloidal particles. The linear relationship between $\ln J$ and $1/\ln(1 + \sigma)$ attained in Fig. 3d allows a good description of the steady-state kinetics by the current 2D nucleation theories, and determination of the average line tension at the interface, γ . The value of γ is $0.27k_B T/a$, in agreement with the average value

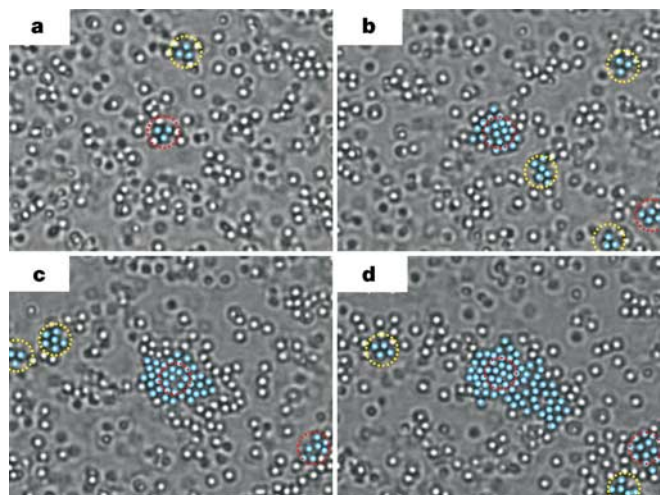


Figure 2 Snapshots of the pre-nucleation process at multiples of $t_0 = 0.294$ s, where t_0 is the timing interval of the images. The radius of the red and yellow circles equals the critical radius r_c at the given supersaturation ($V_{pp} = 2.4$ V, $f = 600$ Hz). The bright dots correspond to the well-focused spheres that are confined on the bottom of the electrodes. The dark dots correspond to the ill-focused spheres, which are suspended in the bulk solution. $t = 14t_0$ (a), $34t_0$ (b), $54t_0$ (c), and $74t_0$ (d).

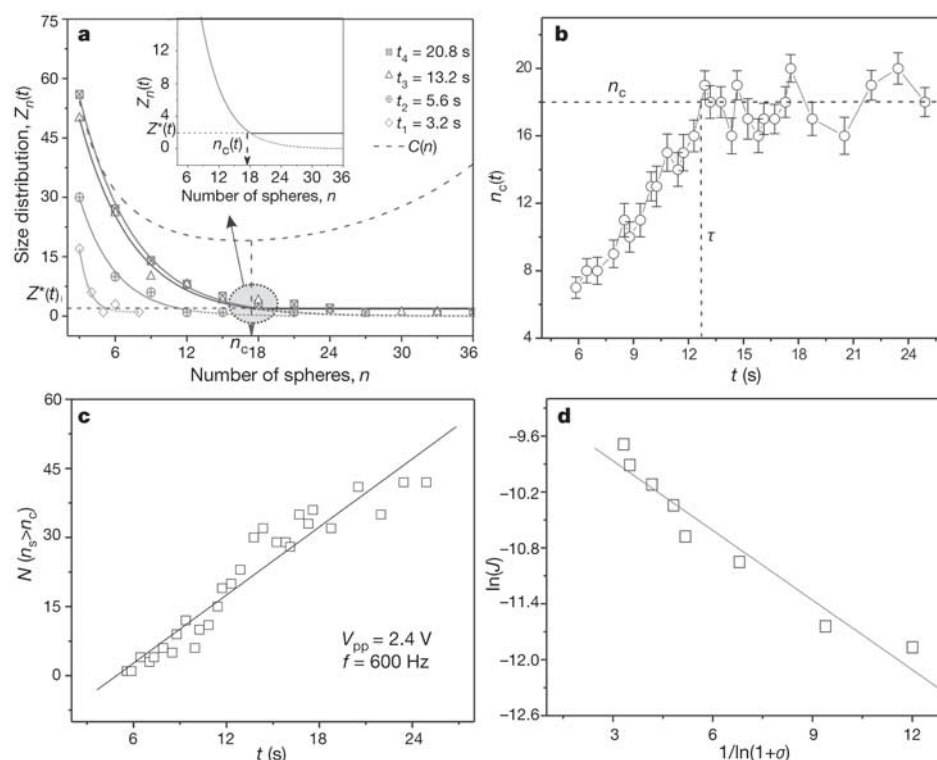


Figure 3 Statistical measurements of parameters of nucleation kinetics. **a**, The typical size distribution of subcritical nuclei at different times ($V_{pp} = 2.4$ V, $f = 600$ Hz), and the equilibrium state distribution based on the Boltzmann distribution. The inset corresponds to $t_4 = 20.8$ s. **b**, $n_c(t)$ versus time. After reaching the induction time τ , $n_c(t)$ becomes stable, thus allowing the determination of the critical size n_c of SSN. Error bars show

standard deviation from the mean. **c**, The number of nuclei having size larger than the critical size at different times. The nucleation rate per unit area can be obtained from the slope of a linear fit to the data points. **d**, The nucleation rate density versus supersaturation (driving force). In the main text we show how the line tension can be calculated from the linear fit to the data points.

$\gamma \approx 0.36 \pm 0.096 k_B T/a$ as calculated from equation (2). Two calculated average values are slightly lower than the numerical result ($\gamma_{\text{line}} \geq 0.43 k_B T/a$) of the hard-sphere system²⁹, the first-order approximation for a real system. This is attributed to the change in the balance of the soft interaction between the electrohydrodynamic attraction and the electrostatic repulsion among the particles^{23,25}.

Our experiments also provide crucial experimental verification for different models and theories^{2–7,27,28,30}. The above data are in very good agreement with classical 2D nucleation. Nevertheless, there is a deviation between the experimental data and the values predicted by the nucleation theorem³⁰ at relatively high driving forces (see Supplementary Figure). Determining the reason for the deviation will require additional investigations.

This unprecedented direct observation of the structure, kinetics and evolution of small crystalline nuclei allows us to characterize the

key processes of nucleation and growth of colloidal crystals, and to determine accurately the important parameters that control these processes. We believe that the present results could lead to a robust technology to produce large and highly ordered colloidal crystals at a relatively low nucleation rate. In Fig. 4, we compare 2D colloidal crystals produced at a constant electric field and at an AEF. At a constant electric field, small and discrete crystallites are formed on the electrode⁸. Using an AEF, the nucleation rate of the 2D crystals can be controlled so as to achieve a high level of perfection in relatively large-area crystals (up to 0.2×0.2 mm). □

Methods

Image recording and processing

In our experiments, we examine quantitatively the AEF-driven nucleation of charge-stabilized polystyrene (PS) spheres (diameter $a = 0.99$ μm , polydispersity $< 5\%$, from Duke Scientific) suspended in deionized water (resistivity ~ 18.2 $\text{M}\Omega \text{ cm}^{-1}$). The aqueous suspension of PS spheres (volume fraction $\phi = 0.2\%$ and $\text{pH} = 4.34$) is sandwiched and sealed in a glass cell composed of two pieces of conductive ITO glass (see experimental set-up, Fig. 1a). The ζ potential of the negative charge of the PS spheres is -26 mV. The continuous *in situ* images are recorded by a digital imaging camera (CoolSNAP cf, Photometrics) which is mounted on an Olympus BX51 microscope. In the reported experiments, we select the appropriate capturing rate (3.4 frames s^{-1}) and resolution ($1,392 \times 1,040$ pixel, corresponding to 258×193 μm under the $50\times$ and long working distance objective) to record the process of nucleation. The sequence of images is simultaneously saved on a hard disk having large volume (120 Gbit) by image processing software (MetaCam). We can identify the individual colloidal spheres and analyse the continuous nucleation process by using codes developed in-house.

Measurement of supersaturation

To measure the supersaturation accurately, we used an ultrasonicator (35 kHz, 10 min) to redisperse the colloidal particles each time before re-applying the AEF. When the concentration on the surface caused the area of the electrode to be fully supersaturated, nuclei appeared and formed. We considered that moment to be the onset of supersaturation, and we measured the concentration of particles on the surface at that instant. When we increased the frequency of the electric field to 3.0 kHz, we noticed that the colloidal crystals were starting to melt, and we measured simultaneously the concentration of the liquid-like particles around the crystallites at the equilibrium point

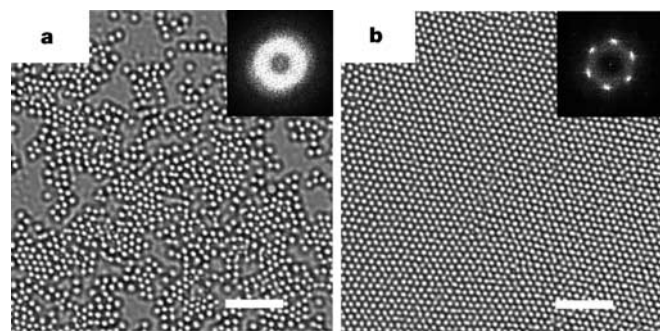


Figure 4 Comparison of the colloidal assemblies obtained under constant electric field and under an AEF. **a**, Constant electric field at 3.2 V. The fast Fourier transform (FFT) in the inset at upper right shows the diffraction ring. **b**, AEF at 2.4 V, 600 Hz. The FFT in the inset at upper right shows the uninterrupted six-fold symmetry of the hexatic phase. Scale bars, 10 μm .

ϕ_m (Fig. 1d). Figure 1d shows a plot of the concentration of the particles that arrived at the surface of the electrode (in terms of the fraction of the occupied area) against the frequency of the applied field. Therefore the supersaturation can be determined from Fig. 1d.

Identification of crystalline nuclei

We used two criteria to identify the crystalline nuclei in the process of nucleation of colloidal particles. (1) Two particles were defined as neighbours if the centre-to-centre distance between them was smaller than a cut-off value of $1.80a$, chosen to be close to the minimum between the first and second peaks of the PCF $g(r)$ for our samples (Fig. 1c). (2) A particle was defined as a member of a 2D crystal if it had more than two neighbours. To exclude the structures of circles or dumbbells (two crystals linked with a long chain) and to consider the structural relaxation, we further confined the bond angles between a particle and its neighbours to $\pi/3 \pm \pi/12$, which corresponds to the deviation from the six-fold symmetric bond angle.

Received 23 November 2003; accepted 10 May 2004; doi:10.1038/nature02630.

1. Sato, K., Furukawa, Y. & Nakajima, K. (eds) *Advances in Crystal Growth* (Elsevier Science, Amsterdam, 2001).
2. Abraham, F. F. *Homogeneous Nucleation Theory* (Academic, New York, 1974).
3. Kashchiev, D. *Nucleation: Basic Theory with Applications* (Butterworth-Heinemann, Oxford, 2000).
4. Laaksonen, A., Talanquer, V. & Oxtoby, D. W. Nucleation: Measurements, theory, and atmospheric applications. *Annu. Rev. Phys. Chem.* **46**, 489–524 (1995).
5. Chernov, A. A. *Modern Crystallography III—Crystal Growth* (Springer, Berlin, 1984).
6. Skripov, V. P. in *Current Topics in Materials Science* Vol. 2 (ed. Kaldis, E.) 328–378 (North-Holland, Amsterdam, 1981).
7. McGraw, R. & Laaksonen, A. Scaling properties of the critical nucleus in classical and molecular-based theories of vapor-liquid nucleation. *Phys. Rev. Lett.* **76**, 2754–2757 (1996).
8. Trau, M., Saville, D. A. & Aksay, I. A. Field-induced layering of colloidal crystals. *Science* **272**, 706–709 (1996).
9. Palberg, T. Crystallization kinetics of repulsive colloidal spheres. *J. Phys. Condens. Matter* **11**, R323–R360 (1999).
10. Larsen, A. E. & Grier, D. G. Like-charge attractions in metastable colloidal crystallites. *Nature* **385**, 230–233 (1997).
11. Yethiraj, A. & van Blaaderen, A. A colloidal model system with an interaction tunable from hard spheres to soft and dipolar. *Nature* **421**, 513–517 (2003).
12. Pusey, P. N. in *Liquids, Freezing and the Glass Transition* (eds Levesque, D., Hansen, J. P. & Zinn-Justin, J.) Ch. 5 (Elsevier, Amsterdam, 1991).
13. Gasser, U., Weeks, E. R., Schofield, A., Pusey, P. N. & Weitz, D. A. Real-space imaging of nucleation and growth in colloidal crystallization. *Science* **292**, 258–262 (2001).
14. Lucadamo, G. & Medlin, D. L. Geometric origin of hexagonal close packing at a grain boundary in gold. *Science* **300**, 1272–1275 (2003).
15. Caccluto, A., Auer, S. & Frenkel, D. Onset of heterogeneous crystal nucleation in colloidal suspensions. *Nature* **428**, 404–406 (2004).
16. Lumsdon, S. O., Kaler, E. W., Williams, J. P. & Velev, O. D. Dielectrophoretic assembly of oriented and switchable two-dimensional photonic crystals. *Appl. Phys. Lett.* **82**, 949–951 (2003).
17. Norris, D. J. & Vlasov, Y. A. Chemical approaches to three-dimensional semiconductor photonic crystals. *Adv. Mater.* **13**, 371–376 (2001).
18. Kumacheva, E., Golding, R. K., Allard, M. & Sargent, E. H. Colloid crystal growth on mesoscopically patterned surfaces: effect of confinement. *Adv. Mater.* **14**, 221–224 (2002).
19. Pan, G. S., Kesavamoorthy, R. & Asher, S. A. Optically nonlinear Bragg diffracting nanosecond optical switches. *Phys. Rev. Lett.* **78**, 3860–3863 (1997).
20. Khopade, A. J. & Caruso, F. Stepwise self-assembled poly(amidoamine) dendrimer and poly(styrenesulfonate) microcapsules as sustained delivery vehicles. *Biomacromolecules* **3**, 1154–1162 (2002).
21. Velev, O. D. & Kaler, E. W. In situ assembly of colloidal particles into miniaturized biosensors. *Langmuir* **15**, 3693–3698 (1999).
22. Fowler, R. & Guggenheim, E. A. *Statistical Thermodynamics* (Cambridge Univ. Press, London, 1960).
23. Yeh, S. R., Seul, M. & Shraiman, B. I. Assembly of ordered colloidal aggregates by electric-field-induced fluid flow. *Nature* **386**, 57–59 (1997).
24. Stiles, P. J. & Regan, H. M. Transient cellular convection in electrically polarized colloidal suspensions. *J. Colloid Interface Sci.* **202**, 562–565 (1998).
25. Trau, M., Sankaran, S., Saville, D. A. & Aksay, I. A. Electric-field-induced pattern formation in colloidal suspensions. *Nature* **374**, 437–439 (1995).
26. Crocker, J. C. & Grier, D. G. Methods of digital video microscopy for colloidal studies. *J. Colloid Interface Sci.* **179**, 298–310 (1996).
27. Liu, X. Y. in *Advances in Crystal Growth Research* (eds Sato, K., Nakajima, K. & Furukawa, Y.) 42–61 (Elsevier Science, Amsterdam, 2001).
28. Liu, X. Y., Maiwa, K. & Tsukamoto, K. Heterogeneous two-dimensional nucleation and growth kinetics. *J. Chem. Phys.* **106**, 1870–1879 (1997).
29. Auer, S. & Frenkel, D. Line tension controls wall-induced crystal nucleation in hard-sphere colloids. *Phys. Rev. Lett.* **91**, 015703 (2003).
30. Bowles, R. K. *et al.* A molecular based derivation of the nucleation theorem. *J. Chem. Phys.* **113**, 4524–4532 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D.W. Li for help with the imaging process, and for discussions. We also thank C. Strom for reading the draft. This work was supported by the Science and Engineering Research Council of Singapore.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to X.Y.L. (phyluxy@nus.edu.sg).

The influence of ridge migration on the magmatic segmentation of mid-ocean ridges

S. M. Carbotte, C. Small & K. Donnelly

Lamont-Doherty Earth Observatory, 61 Rte 9W, Palisades, New York 10964, USA

The Earth's mid-ocean ridges display systematic changes in depth and shape, which subdivide the ridges into discrete spreading segments bounded by transform faults and smaller non-transform offsets of the axis^{1–3}. These morphological changes have been attributed to spatial variations in the supply of magma from the mantle, although the origin of the variations is poorly understood^{1,4,5}. Here we show that magmatic segmentation of ridges with fast and intermediate spreading rates is directly related to the migration velocity of the spreading axis over the mantle. For over 9,500 km of mid-ocean ridge examined, leading ridge segments in the 'hotspot' reference frame coincide with the shallow magmatically robust segments across 86 per cent of all transform faults and 73 per cent of all second-order discontinuities. We attribute this relationship to asymmetric mantle upwelling and melt production due to ridge migration, with focusing of melt towards ridge segments across discontinuities. The model is consistent with variations in crustal structure across discontinuities of the East Pacific Rise, and may explain variations in depth of melting and the distribution of enriched lavas.

Changes in the elevation and cross-axis morphology of a mid-ocean ridge (MOR) often coincide with differences in crustal indicators of magmatic activity, such as the age and composition of axial lavas, presence of crustal magma bodies, and extent of crustal fracturing^{6–8}. These correlations have led to the hypothesis that the morphology of a spreading centre is an indicator of the distribution and delivery of magma from the mantle, with shallow and broad ridge segments believed to reflect enhanced magma supply, whereas deeper, more tectonized segments are magma-starved^{1,9}. Although magmatic segmentation is widely accepted as a fundamental characteristic of a MOR, the significance of this segmentation for the pattern of mantle upwelling is actively debated, particularly for fast-spreading ridges^{4,5,10,11}.

Our analysis of elevation changes along MORs in the context of absolute plate motions reveals a remarkably consistent relationship between magmatic segmentation inferred from ridge morphology and absolute migration velocity of the spreading centre. Plate boundaries migrate at a rate and direction determined by the absolute motions of the bounding plates¹². Currently, all spreading centres are migrating relative to the 'fixed' hotspot reference frame¹³. Along the fast spreading northern East Pacific Rise (NEPR), the spreading centre migrates to the northwest with minor changes in rate and direction along the ridge axis (Fig. 1). Depth profiles along the spreading axis show that shallow ridge segments are consistently offset in the direction of ridge migration relative to the deeper adjacent segments. For the roughly 2,000 km of the NEPR examined (3° to 23°N), this relationship is observed at all transform faults, as well as 6 out of 7 overlapping spreading centres (OSCs) classified as second-order discontinuities (following ref. 1). Analysis of other fast- and intermediate-spreading ridges, encompassing roughly one-fifth of the world's MORs, reveals steps in axial depth consistent with the direction of ridge migration at 86% of transform faults and 73% of second-order discontinuities (Fig. 2, Supplementary Figs 1 and 2).

Steps in ridge elevation across discontinuities, which are ubiqui-

tous along MORs, can not be attributed to cooling related to juxtaposition of older lithosphere across a ridge offset, as this thermal edge effect is expected to affect both segments on either side of a discontinuity equally. Furthermore, thermal edge effects should be smaller at second-order discontinuities where the lithospheric age contrast across the offset is minor (~ 0.1 – 1 Myr). Instead, elevation changes observed at both non-transform and transform offsets span a similar wide range (10 to ~ 500 m), although there is an order of magnitude difference in offset length (typically >50 km for transform faults compared with 5–20 km for non-transform offsets; Fig. 2). Where leading segments are shallower, depth anomalies across discontinuities vary with offset length, with the largest depth anomalies found at offsets of less than ~ 150 km and smaller anomalies at longer transform faults (Fig. 2). In contrast, for the few cases where trailing segments are shallower, depth changes are typically small (<100 m) and there is no trend with ridge offset length.

Why do the leading segments along these MORs overwhelmingly correspond with the shallower, morphologically robust segments? We suggest that these variations in ridge morphology may result from melt focusing across discontinuities from a broad asymmetric zone of mantle upwelling associated with migrating spreading

centres. Kinematic models of passive plate-driven upwelling beneath a migrating MOR predict significant asymmetries in mantle flow paths, with faster mantle upwelling and greater melt production beneath an advancing plate^{14,15}. If these models are correct, differences in the mantle flow field and melt production region across discontinuities are expected as a function of the geometry of ridge offset (Fig. 3).

Seismic and seafloor observations indicate that crustal formation occurs within a narrow zone (1–2 km), and melts within the mantle must be strongly focused to the ridge axis. In the case of a segmented spreading centre, focusing of melts may occur not only from the mantle upwelling zone beneath each ridge segment but also from

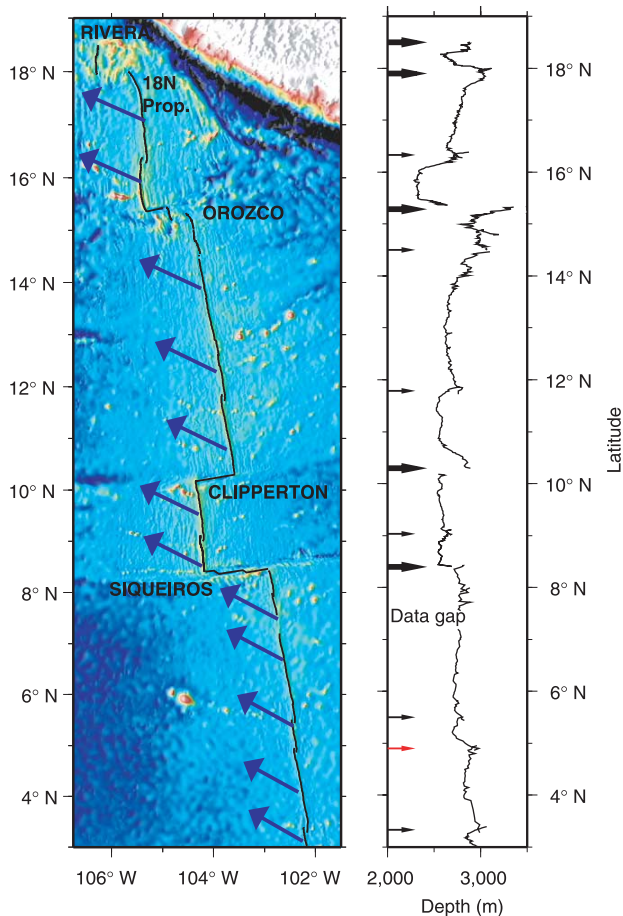


Figure 1 Bathymetry map and axial depth profile along the NEPR. Blue arrows on bathymetry map (left panel) show ridge migration direction calculated from absolute plate motion model HS3-Nuvel1a²⁸. Arrow length is proportional to migration rate ($1 \text{ cm} = 55 \text{ mm yr}^{-1}$). Changes in axial depth (right panel) occur at all first-order discontinuities (thick arrows) and smaller offset second-order discontinuities (thin arrows). With one exception (red arrow), ridge segments offset in the direction of ridge migration are shallower. Differences in ridge elevation can persist for all (for example, south of Siqueiros transform fault) or part (for example, north of Clipperton transform fault) of an adjacent ridge segment. 18N Prop., 18° N Propagator.

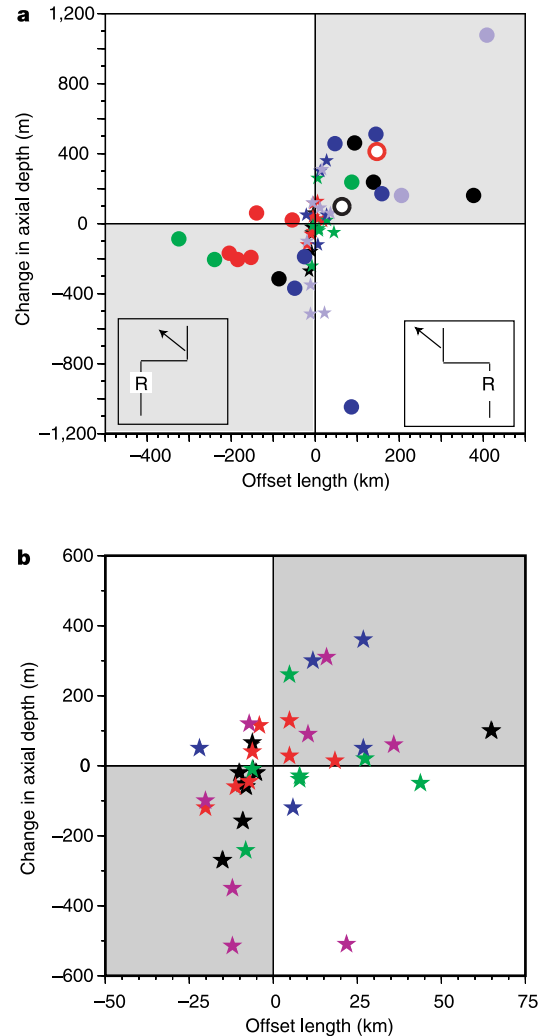


Figure 2 Change in axial depth versus offset length at ridge-axis discontinuities. Change in axial depth is measured as the difference in seafloor depth averaged over a 1-km length of the ridge axis located a uniform distance of 10 km on either side of a discontinuity. Positive axial depth change corresponds to the case where the ridge axis opposite an observer on the spreading axis at R (inset) is shallower. Positive offset length corresponds to the case where the ridge axis steps in the direction of ridge migration relative to observer at R. Grey shading highlights quadrants for which the leading ridge segments are shallower (see text). **a**, Data set includes all transform faults (filled circles), propagating ridge discontinuities (open circles) and non-transform second-order discontinuities (offsets >5 km) along the fast-spreading SEPR from 0°–23° S, 28°–32° S (red), NEPR 3°–23° N (black), Pacific-Antarctic East Pacific Rise 36°–56° S (green), and the intermediate spreading Juan de Fuca-Gorda (purple) and South East Indian ridges 100°–116° E (blue). Bathymetry data used in analysis are from the Ridge Multibeam Synthesis (<http://data.ridge2000.org/bathy>). **b**, Magnified view showing only the smaller-offset second-order discontinuities.

across ridge-axis discontinuities (see, for example, ref. 16). Indeed, melt may be preferentially entrained across discontinuities where horizontal flow paths are shorter to the adjacent ridge axis, leading to extraction of melts at one segment that originated in the upwelling zone of the neighbouring segment. In the case of a migrating spreading centre, melt focusing towards a leading ridge segment will draw primarily from the faster-upwelling, more melt-rich zone of an advancing plate. In contrast, near the ends of trailing segments, the idealized mantle volume that can be tapped for melt lies primarily within the less melt-rich, slower-upwelling zone of a trailing plate (Fig. 3). In this way, trailing segments lie within the wake or shadow zone of a leading segment, potentially entraining a smaller fraction of the melt generated in the upwelling zone of the adjacent leading segment.

The observations suggest that morphological contrasts between adjacent segments can extend for all or part of a ridge segment, with more persistent differences across transform faults (for example, see Fig. 1). The length of this morphological shadow may reflect the volume of melt tapping from an adjacent upwelling zone, although melt redistribution at crustal levels (for example, through dykes or flow within crustal magma bodies) is likely to be important and would obscure such effects.

The strong correlation between ridge migration direction and contrasts in ridge-axis morphology at OSCs along the East Pacific Rise is surprising, given the small scale of these features (Figs 1, 2b). However, with the overlapping geometry of these offsets, the trailing limb of an OSC will lie within the wake of the leading limb along its entire length (Fig. 3). More melt may be entrained (presumably from shallow levels) by the leading OSC limb as a consequence of enhanced melt production beneath advancing plates. Seismic studies of the 9° 03' N OSC¹⁷ suggest that melt at crustal levels is fed to the trailing limb of this OSC from the western, advancing-

plate side of the discontinuity, consistent with this model. An unusually broad westward dipping magma body underlies the trailing limb of the OSC at the northern end of the overlap zone. Further south, small melt channels are imaged along the western edge of this magma lens, suggesting melt feeding from the west.

Our model for the correlation between ridge segment morphology and ridge migration predicts differences in crustal and upper-mantle structure, which can be tested with seismic and gravity studies and explored with numerical experiments. The model requires upper-mantle flow to be asymmetric beneath migrating MORs, and predicts thicker crust at leading segments. The MELT experiment reveals that mantle structure beneath the southern East Pacific Rise (SEPR) from 15° to 19° S is highly asymmetric, with higher inferred melt contents and stronger upper-mantle anisotropy suggesting greater flow-induced strain beneath the advancing Pacific plate^{18–20}. Westward migration of the SEPR has been considered to account for these observations, although existing numerical studies suggest that ridge migration alone is insufficient to explain the large asymmetries observed^{21,22}. New models of asthenospheric flow assuming pressure, temperature, and stress-dependent viscosity show that ridge migration gives rise to small asymmetries in melt production rates of up to 5–11% for intermediate- to fast-spreading MORs²³. Under reasonable assumptions of melt focusing, these models suggest that melt entrainment by adjacent ridge segments from this asymmetric melt production zone could readily account for the topographic changes reported here. These models also predict that depth anomalies across discontinuities should vary with ridge offset length, consistent with the observations.

The only available seismic studies of segment-scale variations in crustal thickness at fast ridges (NEPR 8° 30'–10° 30' N) indicate slightly thinner crust (up to 300 m) at the trailing segment north of the Clipperton transform fault⁸, consistent with our model. The much larger differences in crustal thickness (2–2.5 km) observed along ridge segments south of Clipperton²⁴ may reflect more local-scale processes (for example, the melt source for the Lamont seamounts and propagation history of the OSC). Gravity studies from other fast- and intermediate-spreading regions reveal along-axis variations in mantle Bouguer anomalies consistent with thinner crust (from a few hundred metres to ~1 km) at the trailing segments of both OSCs and transform faults^{10,25,26}. In some of these locations, differences in inferred crustal thickness are much less than predicted

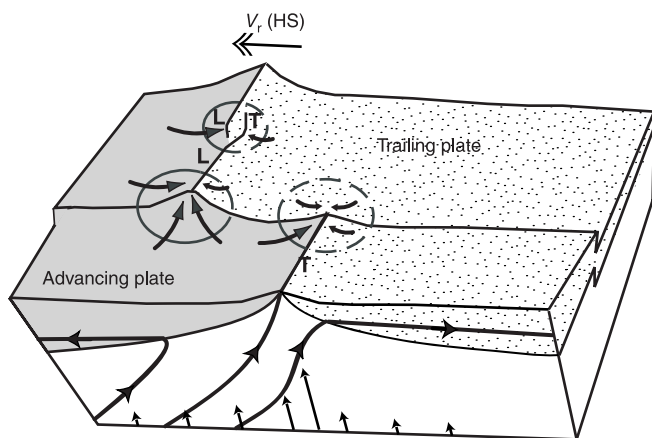


Figure 3 Schematic diagram illustrating asymmetric mantle upwelling beneath a migrating MOR offset at a transform fault and OSC. V_r (HS) indicates the velocity of ridge axis relative to the fixed 'hotspot' reference frame. Following refs 14 and 15, vertical mantle flow is proportional to the slope of the base of the lithosphere and the ridge-normal velocity of each plate with respect to the fixed mantle (vertical arrows). More rapid upwelling and greater melt production is predicted beneath an advancing plate. Stream lines show motion of asthenosphere relative to a stationary observer on the ridge, and indicate asymmetry in the path of mantle material beneath the two plates. The circles encompassing ridge segment ends illustrate idealized regions of mantle melt tapping which include the upwelling zone of the neighbouring segment across the discontinuity (solid/dashed line for advancing/trailing plate respectively, arrows show possible melt migration paths within zone). At leading segments (L), melt can be tapped from the more rapidly upwelling melt-rich zone of an advancing plate across a discontinuity, whereas trailing segments (T) draw melt from the less rapidly upwelling, less-melt-rich trailing plate. At OSCs the same process is envisaged, although probably confined to shallower mantle depths.

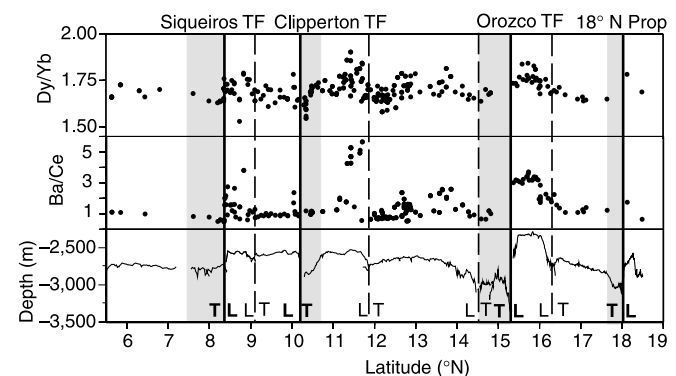


Figure 4 Axial depth and chemical parameters of zero-age lavas along the NEPR. Solid lines labelled in bold font mark large-offset discontinuities; dashed lines correspond to OSCs (offset >5 km). Labels identify leading (L) and trailing (T) segments across discontinuity. Grey shading highlights trailing segments at large-offset discontinuities. For this region, the most enriched and diverse compositions as indicated by incompatible element ratios Ba/Ce are associated with the shallow leading ridge segments. Higher Dy/Yb ratios, which indicate melt originating at greater mantle depths owing to the presence of residual garnet, are also associated with the leading segments. Data from refs 29, 30 and PetDB (<http://petdb.ideo.columbia.edu/petdb/>). TF, transform fault.

for isostatically compensated topography (for example, assuming crustal densities of $2,700 \text{ kg m}^{-3}$, a 300-m difference in crustal thickness would only account for $\sim 100 \text{ m}$ of the depth change across the Clipperton transform fault). Indeed, much of the difference in axial depth across discontinuities may be dynamic owing to thermal isostasy or viscous stresses, plausibly related to asymmetric mantle flow. Determining the relative contribution of dynamic forces versus differences in crustal thickness will require new studies of crustal and upper-mantle structure that target both leading and trailing segments.

The observed correlation between segment morphology and plate kinematics also provides a new framework within which geochemical variations along MORs can be evaluated. Geochemical differences in erupted lavas are observed across discontinuities, and are commonly attributed to melting of a veined mantle. Enriched veins are expected to begin to melt deeper in the mantle²⁷. Our model provides a mechanism whereby these early-melting mantle heterogeneities within the upwelling zone of an adjacent advancing plate could be preferentially entrained beneath leading ridge segments. Along the NEPR, the most enriched and diverse lavas (as indicated by incompatible trace element ratios), as well as those derived from deepest melting, are found along leading segments (Fig. 4). As is typical elsewhere, geochemical studies along this ridge have focused on shallow segments believed to be magmatically robust, and inferences based on the existing data are subject to sampling bias. However, the currently available data for the NEPR are consistent with the notion that geochemical variability along this ridge may reflect the influence of plate kinematics on mantle melt production. Further evaluation of the geochemical implications of our model will require investigations of other areas that include targeted studies of the poorly sampled, trailing segments of MORs. □

Received 22 October 2003; accepted 13 May 2004; doi:10.1038/nature02652.

- Macdonald, K. C. *et al.* A new view of the mid-ocean ridge from the behaviour of ridge-axis discontinuities. *Nature* **335**, 217–225 (1988).
- Lin, J., Purdy, G. M., Schouten, H., Sempere, J.-C. & Zervas, C. Evidence from gravity data for focussed magmatic accretion along the Mid-Atlantic Ridge. *Nature* **344**, 627–632 (1990).
- Cochran, J. R. & Sempere, J.-C. SEIR Scientific Team. The southeast Indian Ridge between 88°E and 118°E : gravity anomalies and crustal accretion at intermediate spreading rates. *J. Geophys. Res.* **102**, 15463–15487 (1997).
- Lin, J. & Phipps Morgan, J. The spreading rate dependence of three-dimensional mid-ocean ridge gravity structure. *Geophys. Res. Lett.* **19**, 13–16 (1992).
- Wang, X., Cochran, J. R. & Barth, G. A. Gravity anomalies, crustal thickness, and the pattern of mantle flow at the fast spreading East Pacific Rise, $9^\circ\text{--}10^\circ\text{N}$; evidence for three-dimensional upwelling. *J. Geophys. Res.* **101**, 17927–17940 (1996).
- Langmuir, C. H., Bender, J. F. & Batiza, R. Petrological and tectonic segmentation of the East Pacific Rise, $5^\circ\text{--}30^\circ\text{N}$. *Nature* **322**, 422–429 (1986).
- Detrick, R. S. *et al.* Multi-channel seismic imaging of a crustal magma chamber along the East Pacific Rise. *Nature* **326**, 35–41 (1987).
- Van Adenonk, H. J. A., Harding, A. J., Orcutt, J. A. & McClain, J. S. Contrast in crustal structure across the Clipperton transform fault from travel time tomography. *J. Geophys. Res.* **106**, 10961–10981 (2001).
- Scheirer, D. S. & Macdonald, K. C. Variation in cross-sectional area of the axial ridge along the East Pacific Rise; evidence for the magmatic budget of a fast spreading center. *J. Geophys. Res.* **98**, 7871–7885 (1993).
- Cormier, M. H., Macdonald, K. C. & Wilson, D. S. A three-dimensional gravity analysis of the East Pacific Rise from 18° to 21°N . *J. Geophys. Res.* **100**, 8063–8082 (1995).
- The MELT Seismic Team. Imaging the deep seismic structure beneath a mid-ocean ridge: The MELT experiment. *Science* **280**, 1215–1218 (1998).
- Stein, S., Melosh, H. J. & Minster, J. B. Ridge migration and asymmetric sea-floor spreading. *Earth Planet. Sci. Lett.* **36**, 51–62 (1977).
- Small, C. & Danyushevsky, L. V. Plate-kinematic explanation for mid-ocean-ridge depth discontinuities. *Geology* **31**, 399–402 (2003).
- Davis, E. E. & Karsten, J. L. On the cause of the asymmetric distribution of seamounts about the Juan de Fuca Ridge; ridge-crest migration over a heterogeneous asthenosphere. *Earth Planet. Sci. Lett.* **79**, 385–396 (1986).
- Schouten, H., Dick, H. J. B. & Klitgord, K. D. Migration of mid-ocean-ridge volcanic segments. *Nature* **326**, 835–839 (1987).
- Madge, L. S. & Sparks, D. W. Three-dimensional mantle upwelling, melt generation, and melt migration beneath segment slow spreading ridges. *J. Geophys. Res.* **102**, 20571–20583 (1997).
- Kent, G. M. *et al.* Evidence from three-dimensional seismic reflectivity images for enhanced melt supply beneath mid-ocean-ridge discontinuities. *Nature* **406**, 614–618 (2000).
- Forsyth, D. W., Webb, S. C., Dorman, L. M. & Shen, Y. Phase velocities of Rayleigh waves in the MELT experiment on the East Pacific Rise. *Science* **280**, 1235–1238 (1998).
- Toomey, D. R., Wilcock, W. S. D., Solomon, S. C., Hammond, W. C. & Orcutt, J. A. Mantle seismic structure beneath the MELT region of the East Pacific Rise from P and S wave tomography. *Science* **280**, 1224–1227 (1998).

- Wolfe, C. J. & Solomon, S. C. Shear-wave splitting and implications for mantle flow beneath the MELT region of the East Pacific Rise. *Science* **280**, 1230–1232 (1998).
- Conder, J. A., Forsyth, D. W. & Parmentier, E. M. Asthenospheric flow and asymmetry of the East Pacific Rise, MELT area. *J. Geophys. Res.* **107** doi:10.1029/2001JB000807 (2002).
- Toomey, D. R. *et al.* Asymmetric mantle dynamics in the MELT region of the East Pacific Rise. *Earth Planet. Sci. Lett.* **200**, 287–295 (2002).
- Katz, R., Spiegelman, M. & Carbotte, S. M. Ridge migration, asthenospheric flow and the origin of magmatic segmentation in the global mid-ocean ridge system. *Geophys. Res. Lett.* (submitted).
- Canales, J. P., Detrick, R. S., Toomey, D. R. & Wilcock, W. S. D. Segment-scale variations in crustal structure of 150- to 300-k.y.-old fast spreading oceanic crust (East Pacific Rise, $8^\circ\text{--}15^\circ\text{N}$) from wide-angle seismic refraction profiles. *Geophys. J. Int.* **152**, 766–794 (2003).
- Weiland, C. M. & Macdonald, K. C. Geophysical study of the East Pacific Rise $15^\circ\text{--}17^\circ\text{N}$; an unusually robust segment. *J. Geophys. Res.* **101**, 20257–20273 (1996).
- Hooft, E. E. & Detrick, R. S. Relationship between axial morphology, crustal thickness, and mantle temperature along the Juan de Fuca and Gorda ridges. *J. Geophys. Res.* **100**, 22499–22508 (1995).
- Hirschmann, M. M. & Stolper, E. W. A possible role for garnet pyroxenite in the origin of the “garnet signature” in MORB. *Contrib. Mineral. Petrol.* **124**, 185–208 (1996).
- Gripp, A. E. & Gordon, R. G. Young tracks of hotspots and current plate velocities. *Geophys. J. Int.* **150**, 321–361 (2002).
- Donnelly, K. E. *The Genesis of E-MORB: Extensions and Limitations of the Hot Spot Model*. Thesis, Columbia Univ. (2002).
- Su, Y.-J. *Mid-ocean Ridge Basalt Trace Element Systematics: Constraints from Database Management, ICPMS Analyses, Global Data Compilation, and Petrologic Modeling*. Thesis, Columbia Univ. (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Katz, M. Spiegelman, R. Buck and W.B.F. Ryan for discussions during the development of these ideas. Special thanks are due to W. Haxby for GeoMapApp, which greatly aided the data analysis. This work was supported by the US National Science Foundation, the Palisades Geophysical Institute, and the Doherty Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.M.C. (carbotte@deo.columbia.edu).

Radiocarbon evidence of mid-Holocene mammoths stranded on an Alaskan Bering Sea island

R. Dale Guthrie

Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska 99775, USA

Island colonization and subsequent dwarfing of Pleistocene proboscideans is one of the more dramatic evolutionary and ecological occurrences^{1–3}, especially in situations where island populations survived end-Pleistocene extinctions whereas those on the nearby mainland did not⁴. For example, Holocene mammoths have been dated from Wrangel Island in northern Russia⁴. In most of these cases, few details are available about the dynamics of how island colonization and extinction occurred. As part of a large radiocarbon dating project of Alaskan mammoth fossils, I addressed this question by including mammoth specimens from Bering Sea islands known to have formed during the end-Pleistocene sea transgression⁵. One date of $7,908 \pm 100 \text{ yr BP}$ (radiocarbon years before present) established the presence of Holocene mammoths on St Paul Island, a first Holocene island record for the Americas. Four lines of evidence—265 accelerator mass spectrometer (AMS) radiocarbon dates from Alaskan mainland mammoths⁶, 13 new dates from Alaskan island mammoths, recent reconstructions of bathymetric plots⁵ and sea transgression rates from the Bering Sea⁵—made it possible to reconstruct how mammoths became stranded in the Pribilof Islands and why this apparently did not happen on other Alaskan Bering Sea islands.

During the first stages of a Holocene sea level rise at 13,000 radiocarbon years before present (referred to hereafter as yr BP), the Pribilofs were simply uplands on a large island lying within a context of dozens of small islands (Fig. 1b). By the time of the St Paul mammoth (*Mammuthus primigenius*) date, 7,900 yr BP (22 m below current sea level), inundation had claimed more land and the uplands that were to become the Pribilofs were separated into several smaller islands quite distant from the mainland (Fig. 1e). Even at that relatively high mid-Holocene sea stand, St Paul Island was approximately 5–10 times larger than today. Some time after 5,000 yr BP (4 m below current sea level) St Paul Island was reduced to near its present size (91 km²) and may have ultimately been too small (Fig. 1f) to support populations of mammoth. Small islands have a limited carrying capacity, and at low numbers inbreeding depression becomes debilitating, then lethal^{7,8}. The absolute size at which an island could no longer support a proboscidean population would depend on productivity of appropriate forage, but judging from empirical evidence the minimum may be over 100 km².

The date obtained from a small, upper third molar (M³) of the St Paul Island mammoth differs markedly in comparison with the date from the last Alaskan mainland mammoth fossil (11,500 ± 150 yr BP; Fig. 2). At that time (and perhaps even as recently as 10,000 yr BP) St Lawrence, St Matthew, Nunivak and Unimak islands were still connected to the mainland (Fig. 1c). But as the sea began to rise the development of Bering Sea islands was not synchronous. We know from dated mammoth bones found on St Lawrence Island that these unglaciated⁹ uplands were used by mammoths during the Last Glacial Maximum (LGM), around 18,000 yr BP when the sea level was 120 m below its current level, but none of the fossils had post-Pleistocene dates (see Supplementary Information). Whatever forces caused mainland mammoth extinction in Alaska and the Yukon Territory (AK–YT) were probably also present in regions that would become islands after 11,500 yr BP. Thus, there is little reason to expect to find Holocene mammoths from St Lawrence and other Bering Sea islands that became isolated after extinction of mammoths from the mainland.

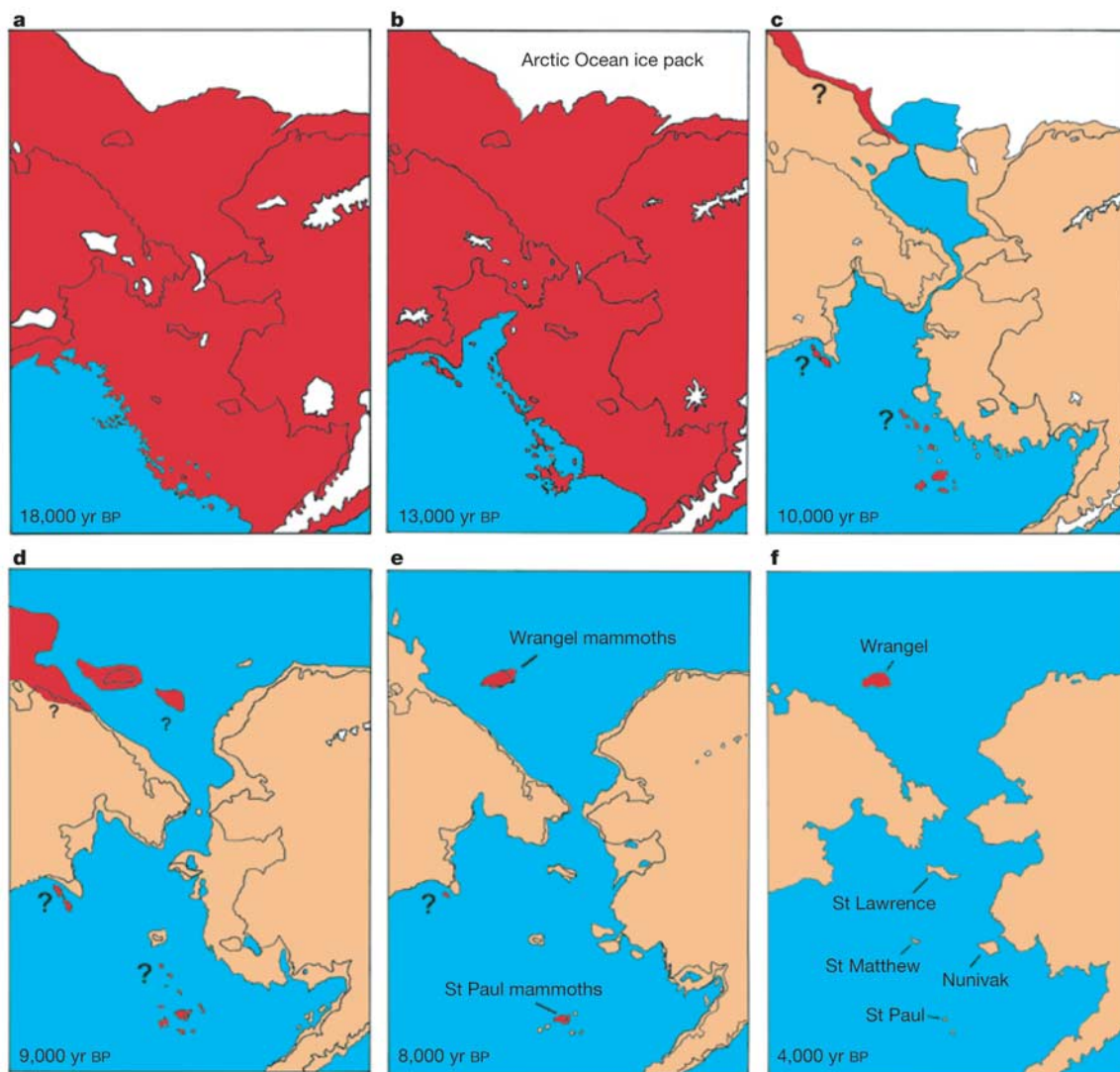


Figure 1 Rise in sea level of the Bering Sea with time, modified from Manley's approximate contours and transgression dates¹⁰. **a–f**, Mammoth distribution (red) at 18,000 (**a**), 13,000 (**b**), 10,000 (**c**), 9,000 (**d**), 8,000 (**e**) and 4,000 (**f**) yr BP is shown. This reconstruction shows how mammoths would have been isolated on the large island containing the Pribilofs around 13,000 yr BP. By 8,000 yr BP the Pribilofs were further

isolated but were much larger in size than today. Note how St Lawrence and St Matthew were still extensions of the mainland until around 10,000 yr BP, well after woolly mammoths are suspected to have become extinct on the AK–YT mainland (see Supplementary Information).

In contrast, the Pribilofs were separated from the mainland possibly as early as 13,000 yr BP (88 m below current sea level), when woolly mammoth were still widespread^{6,10}. In addition to mammoth, caribou (*Rangifer*), bear (*Ursus*) and fox (*Alopex*) fossils have been found on St Paul Island, and are currently under study¹¹.

Sea transgression reconstruction (Fig. 1c, d) makes it clear that other islands existed during the Early Holocene but were eventually inundated. These unnamed islands may have also sustained stranded mammoth populations. Across the Bering Sea, radiocarbon dates show that mammoth populations lasted until after 10,000 yr BP in northernmost Siberia¹⁰. This latter date is germane to the issue of why there is a gap in mammoth dates on Wrangel Island between 12,000 and 8,000 yr BP. Wrangel Island was part of the mainland around 12,000 yr BP but was many kilometres offshore at 8,000 yr BP. Did mammoths somehow recolonize Wrangel Island at 8,000 yr BP or is this gap a taphonomic artefact of what was actually continuous occupation. I leave a question mark (?) on that issue in Fig. 1c.

The productivity on St Paul Island today is rather modest¹² compared with nearby St George Island and the Aleutians. Whereas the latter two are coastal grasslands, St Paul includes some tundra elements^{12,13}. Geological evidence supports the idea that during the Pleistocene this region was quite arid, experiencing loess deposition and reworking. Bathymetric analysis shows that the Yukon River emptied into the Bering Sea just to the south of the Pribilofs during the LGM. Evidence from a pollen core¹³ on Cagalog Lake on St Paul Island shows a dominance of grasses, sedges and *Artemisia* (sage), the ubiquitous spectra typical of cold/arid steppes of the unglaciated north in pre-Holocene times^{14,15}. However, during the Early Holocene (exact dates are disturbed by groundwater movement) on St Paul, the Umbelliferae forbs (probably *Coeloleurum gmelini*) increased in number, suggesting more moisture, a characteristic of Bering Sea rim vegetation today^{12,13}. St Paul Island does not have permafrost, and the season of snow cover is much shorter than its latitudinal equivalent on the mainland. The maritime buffer of the nearby north Pacific and accompanying weather tracts bring unbroken year-round low cloud cover, producing a cool, moist summer with a long growing season in which certain

forbs flourish but trees are excluded.

There are two issues to address: why did mammoths survive later on the islands (St Paul, Wrangel and perhaps other, now inundated, islands) than on the mainland, and why did they fail to survive past a certain point. Proponents of the human overkill theory¹⁶ argue that whereas humans were responsible for the extinction of mammoths on the mainland, offshore island populations were spared because humans had limited access to distant islands until later in the Holocene. Those that see the Late Pleistocene extinctions as the result of ecological changes¹⁷ argue that island mammoths would have been sheltered from the profound climatic and vegetational shifts that favoured the explosion of Holocene large mammal species in competition with Pleistocene ones. However, that controversy remains unresolved.

It is possible that the extinction of island proboscideans may have been a different process than extinction on the mainland. Extinction of mammoths from the California Channel Islands is correlated with the documented arrival of humans around 11,000 yr BP¹⁸. The time of the first human occupation of Wrangel Island is uncertain, but the late timing of mammoth extinction there at around 3,700 yr BP does coincide with the time of marine expansion of northern Denbigh-Arctic small-tool-tradition peoples at about 4,000 yr BP¹⁹. There is archaeological evidence of human occupation of Wrangel Island at least at about 4,300 yr BP²⁰. There is, however, no reason to suspect that humans killed mammoths on St Paul Island; we have no record of human discovery of that island until modern times. Rather, it appears that on St Paul sea transgression separated mammoths from the mainland and a continuation of that sea transgression eventually reduced island areas to the point where life was unsustainable for mammoths.

Closely linked to this matter of Pleistocene island extinction is the controversy over the timing and routes by which humans first entered North America. A 'coastal route' camp²¹ argues for an early colonization (around 13,000 yr BP), whereas the 'overland' camp¹⁸ points to a route between the Cordilleran-Laurentide glaciers, and highlights evidence in the range of 12,000–11,000 yr BP. The St Paul Island mammoth brings a new piece to this puzzle.

If there were coastal watercraft colonists, and they became the

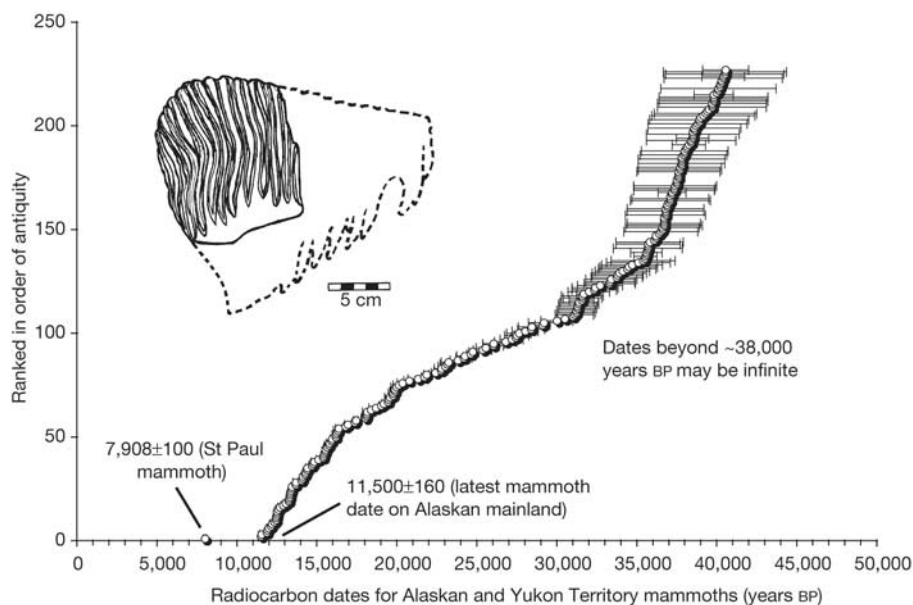


Figure 2 Rank comparison by AMS radiocarbon dates of woolly mammoth from the Alaskan and northwest Canadian mainland showing their relation to the specimen from the Pribilofs. Error bars = 1 σ (see Supplementary Information). The drawing is of the

Pribilof mammoth tooth, with a dashed-line reconstruction of its original contours. Infinite dates beyond 40,500 yr BP are deleted here, but are listed in the Supplementary Information.

continent's 'super-hunters' as per the overkill theory, we might ask why did they not find and kill off mammoths on St Paul? Mammoths on that earlier island complex at 13,000 yr BP would have been easily visible in a treeless landscape when St Paul was separated from the mainland by a narrow channel, in the most likely path of coastal watercraft colonists (Fig. 1b). This becomes part of a broader issue because dwarf mammoths of the California near-shore Channel Islands pose a similar situation. There is no evidence that the Channel Island dwarf mammoths were hunted at around 13,000 yr BP, and they did not become extinct at that time. Rather, these island mammoths became extinct at the time Clovis-aged people invaded the islands at around 11,000 yr BP²¹. Clearly, some island mammoth extinctions were the result of human colonization; however, on St Paul that does not seem to have been the case. □

Methods

Despite ground surveys and excavations on St Matthew Island for potential Quaternary outcrops, and excavated test pits, other researchers and myself have failed to find evidence of Quaternary fossils there. I searched the large collections from Alaska in the University of Alaska Museum, Fairbanks, US National Museum, Washington, and American Museum of Natural History, New York, for fossil bones from the Bering Sea islands. Those specimens are listed in the Supplementary Information. All specimens in this study were dated by the AMS radiocarbon-dating method at the NSF-University of Arizona AMS Facility at Tucson.

The Pribilof M³ was located at the US National Museum (USNM 23455). Provenience data indicate that the tooth was excavated by R. E. Carroll in 1964 from Northeast Point, 75 cm below the surface. The M³ showed a mid-range stage of wear, indicating an adult animal. Maximum width at the 12th plate was 6.5 cm, frequency of enamel plates was 9.23 per 10 cm, and the enamel thickness averaged around 1.2 mm. The roots had been abraded away almost to the ventral bases of the enamel plates, resulting in a smoothly worn contour, suggesting gentle rounding forces like wind-blown sand. The incompleteness of the St Paul specimen makes it difficult to judge body size. Island dwarfing is a general rule for large mammals². Such dwarfing is a process of genetic change, but proboscideans are also known to possess unusual developmental size plasticity in response to conditions during their life^{22,23}. Ongoing work on St Paul may furnish an answer to this question of dwarfing.

I double-checked the first date of 7,908 ± 100 yr BP from the NSF-Arizona Laboratory (AA26010). That split-sample (AA34501) dated at 8,015 ± 85 yr BP. The sample was further re-dated at the Oxford ORAU facility at 8,010 ± 40 yr BP (OxA-13027).

The stable isotope ¹³C-21.5 analysis from the University of Arizona laboratory for the Pribilof M³ fell into the ¹³C mid-range of mainland M³s from that same laboratory that averaged 21.60 ± 0.70 yr (n = 83), so we can assume insignificant marine isotope contamination that might distort the dating by differential marine fractionization of the ¹³C and ¹⁴C isotopes of carbon.

Received 8 March; accepted 27 April 2004; doi:10.1038/nature02612.

- Palombo, M. R. *The World of Elephants*. Proc. 1st Internat. Cong. 486–491 (Comune Di Roma, Consiglio Nazionale della Ricerche, Rome, 2001).
- Sondaar, P. Y. in *Major Patterns of Vertebrate Evolution* (eds Hecht, M. K., Goody, P. C. & Hecht, M. B.) 671–707 (Plenum, New York, 1977).
- Agnew, L. D. Pygmy mammoths, *Mammuthus exilis*, from Channel Islands National Park, California (USA). *Deinsea* 9, 1–39 (2003).
- Vartanyan, S. L., Garutt, V. E. & Sher, A. V. Holocene dwarf mammoths from Wrangel Island in the Siberian Arctic. *Nature* 362, 336–339 (1993).
- Manley, W. F. *Postglacial Flooding of the Bering Land Bridge: A Geospatial Animation v1*. INSTAAR2002 (http://instaar.colorado.edu/QGISL/bering_land_bridge).
- Guthrie, R. D. Rapid body size decline in Alaskan Pleistocene horses before extinction. *Nature* 426, 169–171 (2003).
- Falconer, D. S. *An Introduction to Quantitative Genetics* (Longman, Harlow, 1989).
- Frankham, R. Inbreeding and extinction: island populations. *Conserv. Biol.* 12, 665–675 (2000).
- Hopkins, D. M. & Einarsson, T. Pleistocene glaciation on St. George Is. Pribilof Islands. *Science* 152, 343–345 (1966).
- Sulerzhitsky, L. D. & Romanenko, F. A. The 'twilight' of the mammoth fauna in the Asiatic Arctic. *Ambio* 28, 251–255 (1999).
- Crossen, K. J., Graham, R. W., Veltre, D. W. & Yesner, D. *Abstr. Ann. Geol. Soc. Am. Conf. NW Div.* 424 (Geol. Soc. Am., Seattle, Washington, 2003).
- Scheffer, V. B. Rise and fall of a reindeer herd. *Sci. Monthly* 73, 356–362 (1951).
- Colinvaux, P. A. Historical ecology in Beringia: the south land bridge coast at St. Paul Island. *Quat. Res.* 16, 18–36 (1981).
- Guthrie, R. D. *Frozen Fauna of the Mammoth Steppe* (Univ. Chicago Press, Chicago, 1990).
- Ager, T. A. Late Quaternary vegetation and climate history of the Bering Land Bridge from St. Michael Island western Alaska. *Quat. Res.* 60, 17–31 (2003).
- Martin, P. S. in *Quaternary Extinctions* (eds Martin, P. S. & Klein, R. G.) 345–403 (Univ. Arizona Press, Tucson, 1984).
- Guthrie, R. D. in *Quaternary Extinctions* (eds Martin, P. S. & Klein, R. G.) 259–298 (Univ. Arizona Press, Tucson, 1984).
- Agnew, L. D., Johnson, J. & Morris, D. *Abstr. Am. Quat. Assoc. 17th AMQUA Biennial Meeting* (Am. Quat. Assoc., Univ. Alaska, Anchorage, 2002).
- Fitzhugh, W. W. Global cultural change: new views of circumpolar lands and people. *Anthro. News Natl Mus. Nat. Hist.* 9, 1 (1997).

- Gerasimov, D., Giry, E., Pitulko, V. & Tikhonov, A. *New Materials for the Interpretation of the Site Chertov Ovrug on the Wrangel Island* [in Russian] 379–383 (Mat. 2nd Dikov's conf., Magadan, SVKNII, 2002).
- Fladmark, K. R. Routes: alternate migration corridors for early man in North America. *Am. Antiq.* 4, 55–69 (1977).
- Laws, R. M., Parker, I. S. C. & Johnston, R. C. B. *Elephants and their Habitats: the Ecology of Elephants in North Bunyoro, Uganda* (Clarendon, Oxford, 1975).
- Lister, A. M. Epiphyseal fusion and postcranial age determination in woolly mammoth. *Ann. Nat. Hist. Mus. Rotterdam* 6, 79–87 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements I thank O. Geist and R. Carroll who collected these island fossils, and the curators of the UAM, USNM and ANMH. M. L. Guthrie edited the manuscript. A. Lister and A. J. Stuart provided reviews, and they also corroborated the mammoth date using their NERC Grant. My dating projects have been funded by the NSF.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to R.D.G. (ffrdg@uaf.edu).

Upwelling-driven nearshore hypoxia signals ecosystem and oceanographic changes in the northeast Pacific

Brian A. Grantham^{1*}, Francis Chan^{2*}, Karina J. Nielsen^{4*}, David S. Fox⁵, John A. Barth³, Adriana Huyer³, Jane Lubchenco² & Bruce A. Menge²

¹Washington State Department of Ecology, Coastal and Estuarine Assessment Unit, Olympia, Washington 98504, USA

²Oregon State University, Department of Zoology and

³College of Oceanic and Atmospheric Sciences, Corvallis, Oregon 97331, USA

⁴Sonoma State University, Department of Biology, Rohnert Park, California 94928, USA

⁵Oregon Department of Fish and Wildlife, Marine Resources Program, Newport, Oregon 97365, USA

* These authors contributed equally to this work

Seasonal development of dissolved-oxygen deficits (hypoxia) represents an acute system-level perturbation to ecological dynamics and fishery sustainability in coastal ecosystems around the globe^{1–3}. Whereas anthropogenic nutrient loading has increased the frequency and severity of hypoxia in estuaries and semi-enclosed seas^{3,4}, the occurrence of hypoxia in open-coast upwelling systems reflects ocean conditions that control the delivery of oxygen-poor and nutrient-rich deep water onto continental shelves¹. Upwelling systems support a large proportion of the world's fisheries⁵, therefore understanding the links between changes in ocean climate, upwelling-driven hypoxia and ecological perturbations is critical. Here we report on the unprecedented development of severe inner-shelf (<70 m) hypoxia and resultant mass die-offs of fish and invertebrates within the California Current System. In 2002, cross-shelf transects revealed the development of abnormally low dissolved-oxygen levels as a response to anomalously strong flow of subarctic water into the California Current System. Our findings highlight the sensitivity of inner-shelf ecosystems to variation in ocean conditions, and the potential impacts of climate change on marine communities.

Large-scale fluctuations in ocean climate, such as the El Niño/Southern Oscillation and the Pacific Decadal Oscillation, dominate inter-annual and inter-decadal variability in ocean biology⁶. Iden-

tifying the direct and indirect pathways through which basin-scale changes in ocean conditions act on coastal systems is a central challenge in forecasting the sensitivity and resilience of upwelling ecosystems to climate change. Whereas increases in thermocline depth and upwelling forcing are predicted responses of eastern boundary current systems to climatic warming^{7,8}, the net impacts of regional transport and biogeochemical modifications on the structure and dynamics of coastal ecosystems remain poorly resolved. Assessments of contemporary extremes in ocean conditions offer one direct approach for resolving such uncertainties in ecological sensitivity and response. In summer 2002, anomalous changes in the physical and biogeochemical properties of the California Current System^{9,10} provided an unprecedented opportunity to assess the impacts of large-scale changes in ocean conditions on nearshore upwelling ecosystem dynamics.

Severe inner-shelf hypoxia was detected during coastal oceanographic cruises along the 44.65° N to 44.00° N portion of the California Current System (Figs 1a and 2). Between July and September 2002, bottom dissolved-oxygen concentrations of 0.21–1.57 ml l⁻¹ were found to extend from the shelf break to nearshore stations (2–5 km offshore) along all hydrographic lines (Fig. 2 and Supplementary Fig. 1). Dissolved-oxygen concentrations of <1.43 ml l⁻¹ are generally considered hypoxic¹¹. Inner-shelf cruises within this region along the Strawberry Hill (SH) line recorded bottom dissolved-oxygen concentrations of <1.0 ml l⁻¹

within 700 m of the surf zone, confirming the development of hypoxia as a shelf-wide phenomenon. At the observed height of hypoxia, dissolved-oxygen-deficient bottom waters occupied up to 40 m of the water column (Supplementary Fig. 2a). Transect data from combined cruises indicate that hypoxic conditions covered at least 820 km² of shelf area inshore of the 70-m isobath (Fig. 1a and Supplementary Fig. 1).

The inner-shelf region of the Oregon coast encompasses habitats for several fish populations of important management concern¹². The effects of hypoxia on benthic fish and invertebrate communities were marked. During remotely operated vehicle (ROV) video surveys, only dead fish and dead or moribund invertebrates were found on patch reefs adjacent to the SH line (Fig. 3). In contrast, rockfish densities in previous years averaged 12.8 per 100 m² (Fig. 3) and invertebrate die-offs were not observed in previous surveys. Commercial fishery data revealed crab (*Cancer magister*) mortality to be >75% in crab pots (Fig. 1b) compared with the normal 0%. The Oregon Department of Fish and Wildlife (ODFW) received concurrent reports of large numbers of dead fish and invertebrates washing ashore within the affected region, consistent with our own field observations. Scuba divers reported unusually large aggregations of fish in shallow (<25 m) waters¹³. These unusual observations are consistent with organisms suffocating or seeking oxygen refugia in shallow but atypical habitats.

The development of severe hypoxia in the shallowest reaches of

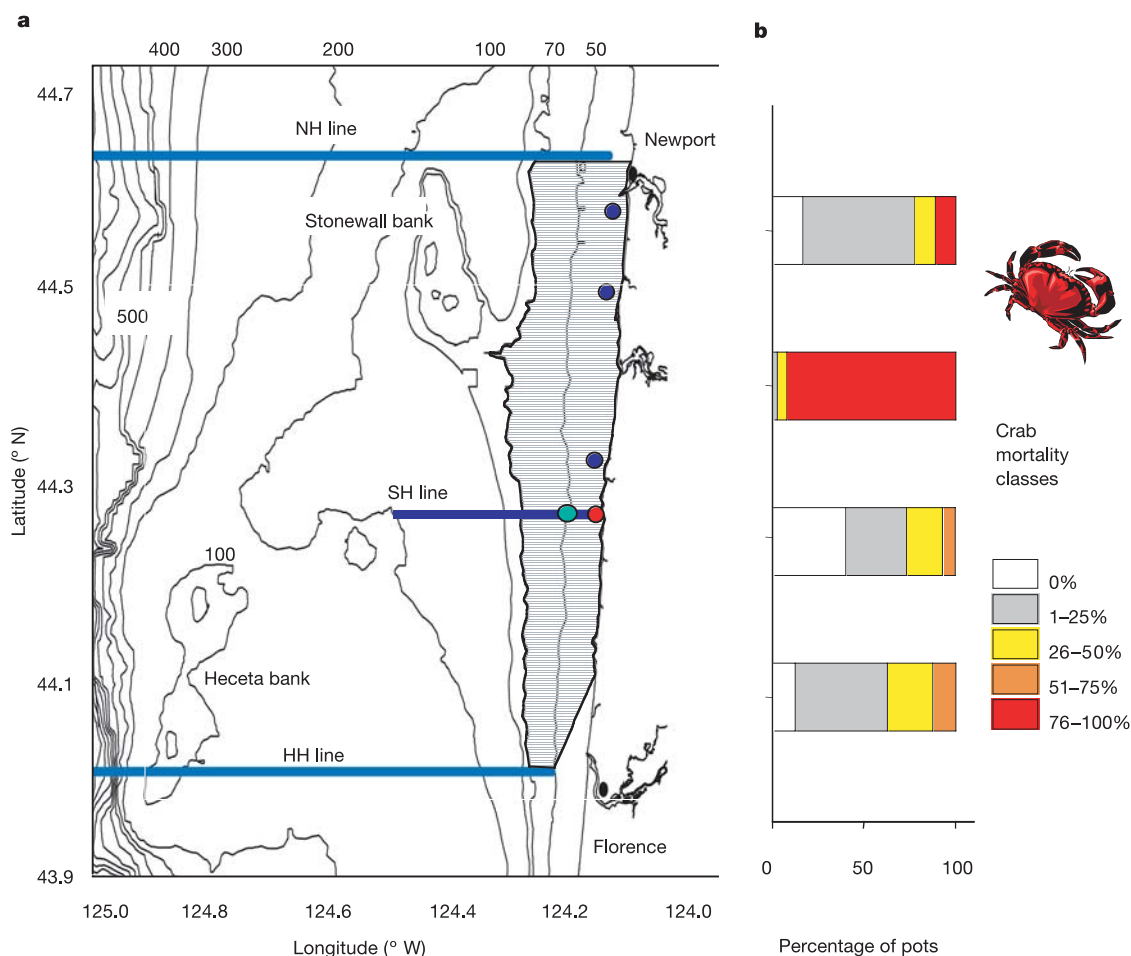


Figure 1 Location of the 2002 hypoxic zone and hydrographic transects off Oregon. **a**, Annual ROV patch reef surveys (green circle). Additional hydrographic stations (blue circles) and the acoustic Doppler current profiler (ADCP) location (red) are indicated. The

minimum estimated spatial extent of the severe hypoxic zone over the inner shelf (~820 km²) is shown (grey). **b**, Proportional mortality of crabs (sampled 16, 17 July 2002) among 80 crab pots deployed within four regions of the hypoxia zone.

the open coast (inshore of the 30-m isobath) is surprising given the potential for air–sea O_2 equilibration during periods of strong turbulent mixing. Significant wave heights reached over 3 m during the hypoxic period. Strong winds favouring upwelling were nevertheless ineffective in eroding stratification on the inner shelf (Fig. 2a, b). Cross-shelf transects revealed the presence of dissolved-oxygen-depleted water offshore on all hydrographic lines (Supplementary Fig. 1). Upwelling forcing thus resulted in the net shoreward transport of cold, saline, dissolved-oxygen-depleted deep water onto the inner shelf (Fig. 2b, c). All bottom dissolved-oxygen values recorded for stations along the 40–55-m isobaths fell below 1.43 ml l^{-1} , suggesting that hypoxia was persistent for at least the 60-day period bracketed by shipborne observations (Fig. 2d). Although upwelling of the dissolved-oxygen-deficient bottom water can create hypoxic conditions over the outer shelf, the development of severe hypoxia inshore of the 70-m isobath is atypical and hypoxia-induced marine life die-offs have not previously been recorded for the Oregon shelf. The anomalous nature of the 2002 hypoxic event is strongly evident when compared with recent and historical records (see Methods). Shelf stations along the Newport Hydrographic line exhibited bottom dissolved-oxygen values that were sharply depressed relative to upwelling season means (Fig. 4a, b).

Summertime upwelling induces the transport of deep-sea water that is low in dissolved oxygen onto continental shelves, and results in the intersection of nutrient-rich halocline water with the sea

surface¹⁴. Consequently, the presence of bottom water that is anomalously low in dissolved oxygen across the shelf can reflect decreases in the oxygen content of upwelled source water before its arrival on the shelf, and/or increases in shelf export production in 2002. An extensive oxygen minimum zone (OMZ), where dissolved oxygen falls below 0.5 ml l^{-1} , exists along the continental margins of the northeast Pacific Ocean¹⁵. Off the coast of Oregon the OMZ is usually located off the shelf at depths of 700–900 m (refs 15, 16). The source of water (with salinity of 33.9) that is upwelled into the bottom boundary layer of the inner shelf lies between ~100–200 m (Supplementary Fig. 2), well above the observed upper boundary of the OMZ. Water that has upwelled from these depths typically has dissolved-oxygen concentrations of $1.8\text{--}3.6 \text{ ml l}^{-1}$ (ref. 16); however, in 2002, dissolved oxygen in source water along the 33.9 isohaline at the shelf break fell to near or below hypoxic levels ($1.27\text{--}1.67 \text{ ml l}^{-1}$) (Fig. 4c).

Although dissolved oxygen had already reached hypoxic levels at the shelf break, respiration can further exacerbate dissolved-oxygen deficits as bottom boundary water transits shoreward over the shelf, particularly if surface organic carbon sources are sizeable. During summer upwelling a coastal jet of the southward-flowing California Current follows the shelf break offshore around Heceta and Stonewall banks^{14,17}, with weak mean flows inshore of the two banks over the shelf¹⁸. Chlorophyll *a* concentrations in this region are consistently higher than those in surrounding areas and are fuelled by outcropping of nutrient-rich water (salinity, 32.8–33.8) from the

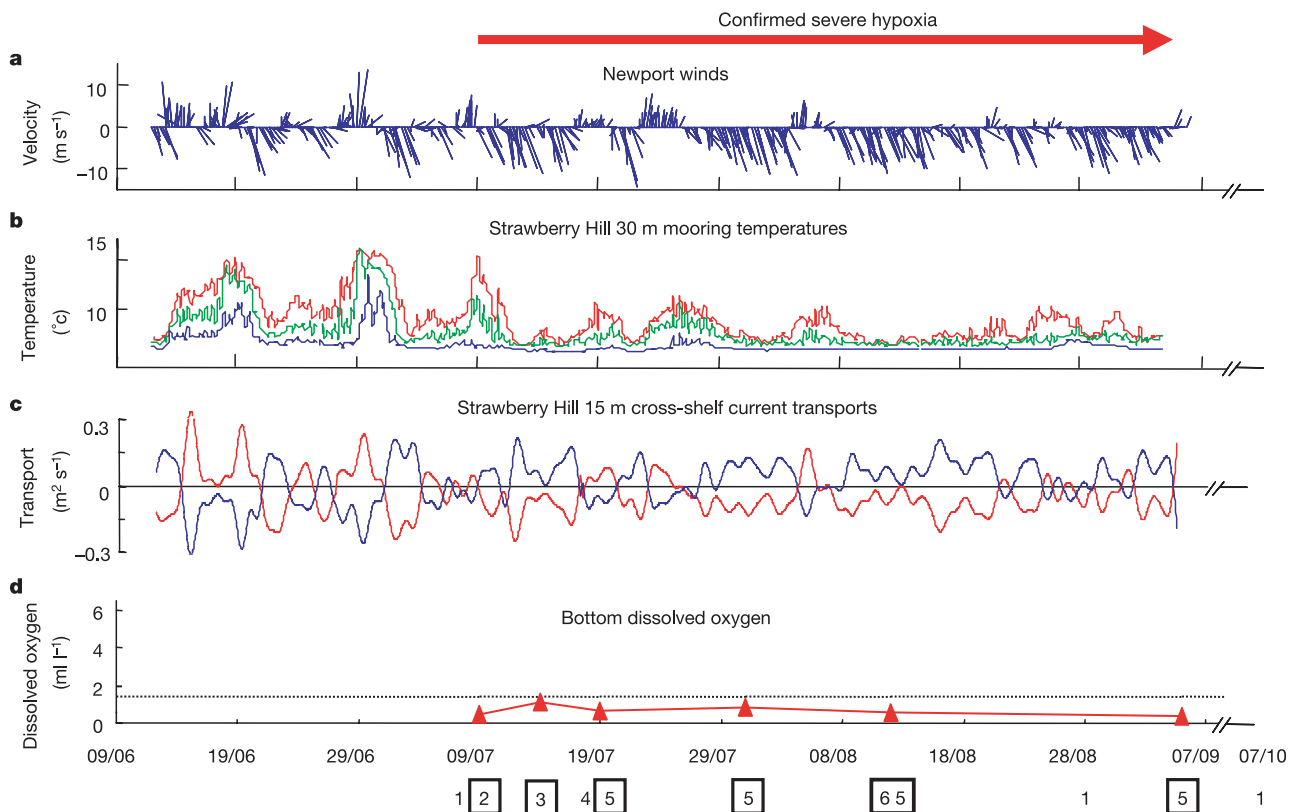


Figure 2 Timeline of observations, both continuous (**a–c**) and discrete (1–6). Numbers 1–6 below the x axis are positioned along the timeline at the times when the corresponding surveys/station measurements were conducted. Boxed numbers relate to the red triangles in **d** that are directly above. 1, ODFW survey; 2, GLOBEC NH line; 3, GLOBEC HH line; 4, crab pot retrieval; 5, PISCO SH line; 6, GLOBEC SH line. **a**, Wind velocities. Negative values represent upwelling-favourable winds. **b**, Temperatures at

depths of 4 m (red), 14 m (green) and 29 m (blue). **c**, ADCP-derived cross-shelf transport estimates for surface wind-driven (red) and bottom boundary (blue) layers. Positive values represent onshore flows. **d**, Bottom dissolved-oxygen concentration (dotted line marks hypoxia threshold) measured at the indicated inner-shelf (40–55 m) stations along the NH, SH and HH lines.

halocline (Supplementary Fig. 3). In 2002, halocline source water was 1 °C colder and nitrate concentrations were elevated by 11.6 μM (or 64%) over past years^{9,10}. As a result, standing stocks of phytoplankton across the bank were two to three times higher than those observed in the four preceding years¹⁹ and surface dissolved oxygen was strongly elevated relative to long-term means (Fig. 4a, b). Inner-shelf chlorophyll *a* concentrations were also high throughout the study (Supplementary Table 1), with maximum concentrations of 219 $\mu\text{g l}^{-1}$ occurring during a massive *Thalassiosira* bloom in mid-August (Supplementary Fig. 2). Dissolved-oxygen concentrations along the 33.9 isohaline decreased from the shelf break to the inner shelf (Supplementary Fig. 1), suggesting that a further depletion of oxygen occurred over the shelf as a result of respiration of fluxed carbon.

Hypoxia is well documented for other major eastern boundary currents, where its frequency and intensity have been linked to

inter-annual changes in ocean climate^{11,20,21}. It now seems that large-scale anomalies in northeast Pacific conditions have given rise to the novel emergence of severe inner-shelf hypoxia in Oregon. In July 2002, the permanent halocline that is present off the coasts of Oregon and Vancouver Island was 1 °C colder, and surface salinities were lower, than in previous years¹⁰. Data on sea-surface height²², drifter tracks²³ and mid-shelf currents²⁴ suggest an anomalous invasion of nutrient-rich, subarctic water into the California Current System as being the causative agent behind these changes¹⁰. The observed deviations in the circulation of the California Current System further reflect large-scale wind stress anomalies present over the northeast Pacific in 2002 (ref. 25). Because upwelling transports waters that can be both dissolved-oxygen deficient and nutrient rich onto productive continental-shelf and seamount systems, hypoxia may represent a general and critical link between climatic variability, shifts in ocean circulation and marine ecological change. The

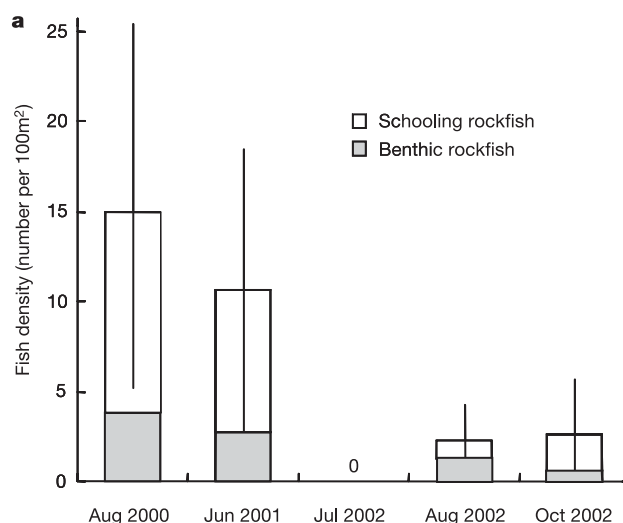


Figure 3 Impact of hypoxia on rockfish communities. **a**, Densities of schooling and benthic rockfish (*Sebastes* spp.) near the SH line from annual ROV surveys in 2000–02. Both schooling and benthic rockfish densities differed significantly between years (analysis of variance: $F = 7.75$, degrees of freedom (d.f.) = 2, 10, $P = 0.009$; $F = 4.71$, d.f. = 2, 10, $P = 0.036$, respectively). Schooling rockfish were lower in October 2002 than in 2000 or 2001 (Scheffe test: $P = 0.013$ and $P = 0.041$, respectively). Benthic rockfish densities were lower in October 2002 than in 2000

(Scheffe test: $P = 0.042$), but did not differ from 2001 (Scheffe test: $P = 0.161$). Error bars indicate 95% confidence intervals. Rockfish species observed included: black (*Sebastes melanops*), canary (*S. pinniger*), copper (*S. caurinus*), quillback (*S. maliger*), yelloweye (*S. ruberrimus*) and yellowtail (*S. flavidus*) rockfish as well as unidentified juveniles and lingcod (*Ophiodon elongatus*). Lingcod, yelloweye and canary rockfish are classified as 'over-fished' by the National Marine Fisheries Service. **b**, **c**, ROV survey images of rockfish, August 2000 (**b**) and mortality, July 2002 (**c**).

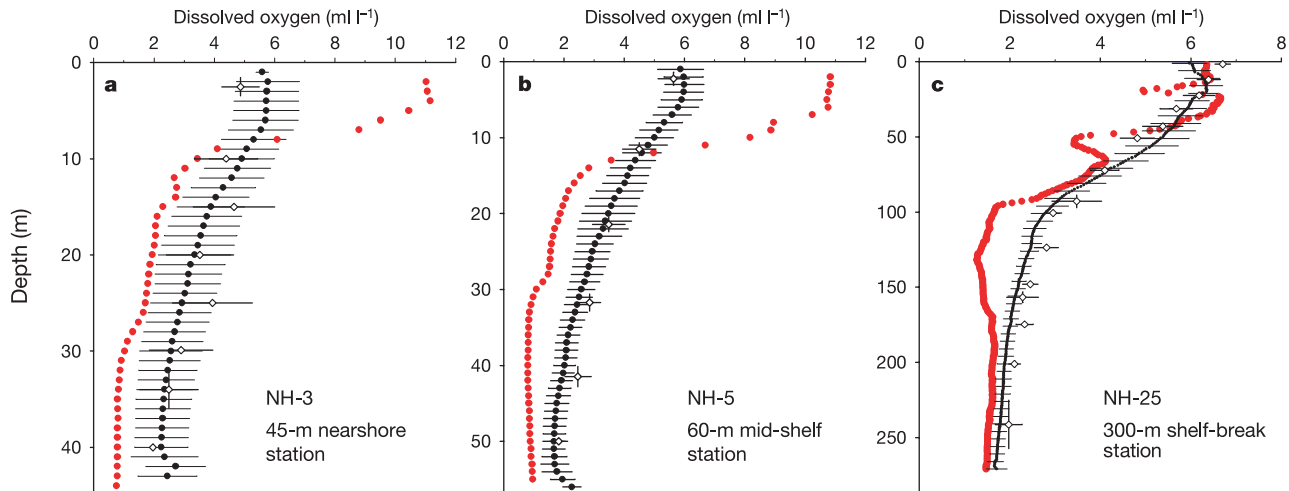


Figure 4 Dissolved-oxygen profiles. **a–c**, July 2002 (red circles), recent (black circles) and historical (open diamonds) summertime (July to September) mean ($\pm 95\%$ confidence intervals) dissolved-oxygen profiles for the nearshore (**a**), mid-shelf (**b**) and

shelf-break (**c**) stations. Recent values span 1998–2001. Historical values cover: **a**, 1967–69, 1972; **b**, 1960–69, 1972; **c**, 1960–69, 1972. NH-3, NH-5 and NH-25 refer to stations along the Newport Hydrographic line.

abrupt emergence of severe shallow-water hypoxia in the California Current System further highlights the importance of ecological thresholds in mediating ecosystem sensitivity to climatic variability, and reinforces the fundamental need for ecosystem-based management of coastal fisheries and habitats. □

Methods

Remotely operated vehicle surveys

ODFW's Marine Resources Program began conducting annual surveys of rockfish and rocky reef habitat along the Oregon coast in 2000. The ROV travels ~1 m above the sea floor and records information on fish and habitat to an average maximum distance of 4.5 m out from the vehicle, using a video camera and laser-based measurement system. In July 2002, six strip transects covering a total area of ~250 m² were surveyed. Data reports and detailed methods are available online²⁶.

Hydrographic transects

Since July 1997, the GLOBEC LTOP program has conducted three to five surveys annually along the Newport Hydrographic (NH) and Heceta Head (HH) lines (44.65 N and 44.0 N, respectively). Dissolved oxygen has been measured since August 1998. A GLOBEC SeaSoar²⁷ survey of the Heceta/Stonewall bank region was conducted from 9–11 July 2002 (Supplementary Fig. 1). The SeaSoar is a towed, undulating platform equipped with a SeaBird SBE9/11 plus conductivity–temperature–depth (CTD) profiler and a WETLabs WETstar fluorometer. On 11 August 2002, temperature, conductivity, depth, dissolved oxygen and chlorophyll *a* fluorescence were sampled at five stations (40–100 m in depth) along the SH line using a SeaBird SBE9/11 plus CTD and a SBE43 dissolved-oxygen sensor, and a Seapoint fluorometer. LTOP and SeaSoar data reports, including methods, are available online^{16,27}. Additional historical-dissolved-oxygen data for NH line stations were obtained from the World Ocean Database (<http://www.nodc.noaa.gov>).

Temperature, conductivity and chlorophyll *a* fluorescence were sampled along the SH line and additional nearshore stations on 19 and 31 July, 12 August and 5 September 2002 using a SeaBird SBE25 CTD with a WETLabs WETstar fluorometer. Chlorophyll *a* and dissolved-oxygen concentrations were obtained from Niskin bottle samples. Dissolved oxygen was analysed immediately with a YSI 556 polarographic oxygen meter that was calibrated in subsequent cruises against Winkler titration-based measurements. Chlorophyll *a* samples were filtered on board the research vessel, packed on ice and returned to the laboratory for analysis on a Turner TD-700 fluorometer.

Mooring-based measurements

Inner-shelf moorings were deployed at 15- and 30-m depths along the SH line. The 15-m mooring was equipped with a SeaBird SBE 16 conductivity and temperature module. Both moorings were outfitted with thermistors (Onset StowAway XTI) that spanned the water column. Estimates of the cross-shelf surface and bottom transports were found using a bottom-mounted ADCP (RD Instruments 600 kHz Workhorse Sentinel) at 15 m along the SH line. Measurements were rotated into their principal axis (4° true), and depth-averaged mean cross-shelf velocities were first subtracted from each velocity profile. The resulting zero-mean profiles were extrapolated vertically to the surface and the bottom, giving a profile that covers the entire water column. The surface (bottom) transport is then defined and computed as the integral of this zero-mean velocity profile from the surface (bottom) down (up) to the first zero crossing. Hourly average transport values were low-pass filtered with a 40-h cutoff to remove tidal variations.

Crab mortality data

A commercial fisherman recorded the GPS (global positioning system) coordinates and mortality rates of dungeness crabs (*C. magister*) retrieved from 80 crab pots set inshore of the Heceta/Stonewall bank area. The traps were set at depths of 8–50 m, allowed to soak for 7–10 days and retrieved on 16–17 July 2002. The average number of crabs per pot for each region (Fig. 1) was 21 (± 3 s.d.).

Received 12 March; accepted 27 April 2004; doi:10.1038/nature02605.

1. Naqvi, S. W. A. *et al.* Increased marine production of N₂O due to intensifying anoxia on the Indian continental shelf. *Nature* **408**, 346–349 (2000).
2. Diaz, R. J. Overview of hypoxia around the world. *J. Environ. Qual.* **30**, 275–281 (2001).
3. Rabalais, N. N. & Turner, R. E. in *Coastal Hypoxia: Consequences for Living Resources and Ecosystems* (eds Rabalais, N. N. & Turner, R. E.) 1–36 (American Geophysical Union, Washington DC, 2001).
4. National Research Council. *Coastal Waters: Understanding and Reducing the Effects of Nutrient Pollution* (National Academy Press, Washington DC, 2000).
5. Pauly, D. & Christensen, V. Primary production required to sustain global fisheries. *Nature* **374**, 255–257 (1995).
6. Chavez, F. P., Ryan, J., Lluch-Cota, S. E. & Niquen, M. From anchovies to sardines and back: Multidecadal change in the Pacific Ocean. *Science* **299**, 217–221 (2003).
7. Bakun, A. Global climate change and intensification of coastal ocean upwelling. *Science* **247**, 198–201 (1990).
8. Roemmich, D. & McGowan, J. Climatic warming and the decline of zooplankton in the California Current. *Science* **267**, 1324–1326 (1995).
9. Wheeler, P. A., Huyer, A. & Fleischbein, J. Cold halocline, increased nutrients and higher productivity off Oregon in 2002. *Geophys. Res. Lett.* [online] **30**, 8021 (2003) (doi:10.1029/2003GL017395).
10. Freeland, H. J., Gatién, G., Huyer, A. & Smith, R. L. Cold halocline in the northern California current: an invasion of subarctic water. *Geophys. Res. Lett.* [online] **30**, 1141 (2003) (doi:10.1029/2002GL016663).
11. Diaz, R. J. & Rosenberg, R. Marine benthic hypoxia: a review of its ecological effects and the behavioural responses of benthic macrofauna. *Oceanogr. Mar. Biol. Ann. Rev.* **33**, 245–303 (1995).
12. National Marine Fisheries Service. *Our Living Oceans* (NOAA Technical Memo NMFS-F/SPO-41, US Department of Commerce, Washington DC, 1999).
13. Dillman, T. *Abundance of Fish on Area Reefs Delights, Puzzles Divers* (Newport News-Times, Newport, Oregon, 2002).
14. Huyer, A. Coastal upwelling in the California current system. *Prog. Oceanogr.* **112**, 259–284 (1983).
15. Kamykowski, D. & Zentara, S. J. Hypoxia in the world ocean as recorded in the historical data set. *Deep-Sea Res.* **37**, 1861–1874 (1990).
16. Fleischbein, J. Northeast Pacific long term observation program. *US GLOBEC* (<http://ltop.coas.oregonstate.edu/~ctd/>) (2003).
17. Huyer, A., Smith, R. L. & Pillsbury, R. D. Observations in a coastal upwelling region during a period of variable winds (Oregon coast, July 1972). *Tethys* **6**, 391–404 (1974).
18. Oke, P. R. *et al.* A modelling study of the three-dimensional continental shelf circulation off Oregon. Part I: Model-data comparisons. *J. Phys. Oceanogr.* **32**, 1360–1382 (2002).
19. Thomas, A. C., Strub, T., Brickley, P. & James, C. Anomalous satellite-measured chlorophyll concentrations in the northern California Current in 2001–2002. *Geophys. Res. Lett.* [online] **30**, 8022 (2003) (doi:10.1029/2003GL017409).
20. Bailey, G. W., Beyers, C. J. B. & Lipschitz, S. R. Seasonal variation of oxygen deficiency in waters off southern south west Africa in 1975 and 1976 and its relation to the catchability and distribution of the cape rock lobster *Jasus lalandii*. *S. Afr. J. Mar. Sci.* **3**, 197–214 (1985).
21. Morales, C. E., Hormazabal, S. E. & Blanco, J. L. Interannual variability in the mesoscale distribution of the depth of the upper boundary of the oxygen minimum layer off northern Chile (18–24S): Implications for the pelagic system and biogeochemical cycling. *J. Mar. Res.* **57**, 909–932 (1999).
22. Strub, P. T. & James, C. Altimeter estimates of anomalous upwelling along the northern California Current during 2000–2002. *Geophys. Res. Lett.* [online] **30**, 8025 (2003) (doi:10.1029/2003GL017513).
23. Barth, J. A. Anomalous southward advection during 2002 in the northern California Current:

- evidence from Lagrangian surface drifters. *Geophys. Res. Lett.* [online] **30**, 8024 (2003) (doi:10.1029/2003GL017511).
24. Kosro, P. M. Enhanced southward flow over the Oregon shelf in 2002: a conduit for subarctic water. *Geophys. Res. Lett.* [online] **30**, 8023 (2003) (doi:10.1029/2003GL017436).
25. Murphree, T., Bograd, S. J., Schwing, F. B. & Ford, B. Large-scale atmosphere–ocean anomalies in the northeast Pacific during 2002. *Geophys. Res. Lett.* [online] **30**, 8026 (2003) (doi:10.1029/2003GL017303).
26. Fox, D. S., Amend, M., Merems, A. & Appy, M. Near-shore rocky reef assessment. *Oregon Dept. Fish Wildl. Mar. Prog.* (http://hmsc.oregonstate.edu/odfw/habitat/reports/2000kelp_reef_report.pdf) (2000).
27. O'Malley, R. SeaSoar observations during the Coastal Ocean Advances in Shelf Transport (COAST) mesoscale surveys I and II. (<http://damp.coas.oregonstate.edu/coast/seasoar>) (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Kirincich for ADCP analyses, S. Oda and A. Walters for assistance and A. Pazar for crab fishery data. The Partnership for Interdisciplinary Studies of Coastal Oceans, funded by the David and Lucile Packard Foundation, and the US GLOBEC program, jointly funded by NSF and NOAA, both contributed to this work. J.A.B. acknowledges additional funding from NSF. ODFW ROV survey work was funded in part by the Oregon Department of Land Conservation and Development through NOAA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.C. (chanft@science.oregonstate.edu).

Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene

Miranda M. Lim¹, Zuoxin Wang², Daniel E. Olazábal¹, Xianghui Ren³, Ernest F. Terwilliger³ & Larry J. Young¹

¹Center for Behavioral Neuroscience and Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia 30322, USA

²Department of Psychology and Neuroscience Program, Florida State University, Tallahassee, Florida 32306, USA

³Harvard Institutes of Medicine and Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, USA

The molecular mechanisms underlying the evolution of complex behaviour are poorly understood. The mammalian genus *Microtus* provides an excellent model for investigating the evolution of social behaviour. Prairie voles (*Microtus ochrogaster*) exhibit a monogamous social structure in nature, whereas closely related meadow voles (*Microtus pennsylvanicus*) are solitary and polygamous¹. In male prairie voles, both vasopressin and dopamine act in the ventral forebrain to regulate selective affiliation between adult mates, known as pair bond formation, as assessed by partner preference in the laboratory^{2–4}. The vasopressin V1a receptor (V1aR) is expressed at higher levels in the ventral forebrain of monogamous than in promiscuous vole species⁵, whereas dopamine receptor distribution is relatively conserved between species. Here we substantially increase partner preference formation in the socially promiscuous meadow vole by using viral vector V1aR gene transfer into the ventral forebrain. We show that a change in the expression of a single gene in the larger context of pre-existing genetic and neural circuits can profoundly alter social behaviour, providing a potential molecular mechanism for the rapid evolution of complex social behaviour.

Fewer than 5% of mammalian species have a monogamous social structure, which typically includes pair bond formation between

adult mates and the biparental care of offspring⁶. Central administration of vasopressin facilitates each of these monogamous-typical behaviours in male prairie voles through V1aRs in the brain^{7–9}. The distribution of V1aRs in the brain varies considerably between monogamous and promiscuous vole species⁵ (Fig. 1).

Species-specific patterns of V1aR expression have provided insight into the neural mechanisms underlying pair bond formation. In particular, the ventral pallidum, located within the ventral forebrain and the mesolimbic dopamine reward pathway, highly expresses V1aRs in monogamous prairie and pine voles (*Microtus pinetorum*), but not in promiscuous meadow or montane voles (*Microtus montanus*)^{5,10}. Site-specific infusion of a selective V1aR antagonist into the ventral pallidum blocks pair bond formation in prairie voles⁴. Ventral forebrain V1aR expression is also higher in both the monogamous *Peromyscus* California mouse and the monogamous marmoset monkey than in promiscuous *Peromyscus* or primate species^{11–14}. Thus, V1aRs in the ventral forebrain are crucial for pair bond formation, and this V1aR pattern seems to be correlated with monogamous social organization across diverse taxa.

Although genetic analysis of V1aR shows more than 99% conservation of gene sequence between vole species, monogamous prairie and pine voles have an expansion of repetitive microsatellite DNA in the 5' regulatory region of the gene, whereas promiscuous montane and meadow voles do not⁹. Furthermore, species-specific V1aR expression patterns seem to be determined by proximate regulatory sequences of the gene, because mice transgenic for the prairie vole V1aR coding sequence and its flanking regions display prairie-like patterns of V1aR binding in the brain⁹. Because microsatellite DNA is highly unstable, it is possible that instability in the V1aR regulatory region could result in altered V1aR expression in different brain regions, leading to differences in social behaviour. We proposed that increased V1aRs in the ventral pallidum, in particular, could shift some individuals within a species to form pair bonds and ultimately result in the stable selection of monogamous social organization. To test this hypothesis directly, we used viral vector-mediated gene transfer to overexpress V1aR in the ventral pallidum in the socially promiscuous meadow vole, in essence recreating a singular evolutionary event in the laboratory.

The study consisted of three groups of meadow voles. The experimental group (V1aR-vp) overexpressed V1aR bilaterally in the ventral pallidum ($n = 11$). The first control group (Ctrl-vp) received ventral pallidal infusions of a vector expressing the *lacZ* gene ($n = 11$). The second control group (Ctrl-other) consisted of animals whose viral injections were inadvertently placed outside the ventral pallidum ($n = 9$); these animals were regrouped, *ex post facto*, in the analysis after the completion of behavioural testing. V1aR autoradiography revealed a significant elevation of V1aR binding in the ventral pallidum of the V1aR-vp animals to about threefold that in the two control groups (Fig. 2). This is comparable to the degree of V1aR binding observed in the prairie vole (Fig. 1e).

All animals were paired with a behaviourally receptive female for 24 h, and subsequently placed in a partner preference test in which the animal could access both the partner and a novel female of comparable stimulus value. During the 3-h test, the time spent in side-by-side contact ('huddling') with each female was recorded.

The V1aR-vp group spent significantly more time huddling with the partner than the stranger ($P < 0.01$, Student's *t*-test), whereas the Ctrl-vp and Ctrl-other groups did not ($P > 0.05$, Student's *t*-test) (Fig. 3a). In addition, all the animals in the V1aR-vp group spent more time huddling with the partner than the stranger (11 of 11), whereas control animals were uniformly distributed across a wide range and did not prefer the partner significantly more than expected by random chance (12 of 20; $P > 0.05$, χ^2 test) (Fig. 3b). V1aR-vp animals also spent significantly more time in total side-by-side contact with the partner than the control animals ($P < 0.01$,

Student's *t*-test), but not significantly more time with the stranger than the control animals ($P > 0.05$, Student's *t*-test), indicating that their partner preference was not a result of a generalized, non-selective increase in affiliative behaviour.

Although several control animals did seem to display selective affiliation with the mate, none of them were placed within the top five strongest pair bonds, as defined by the total time spent huddling with the partner. However, this illustrates that individual variation does exist within the meadow vole population, indicating that the potential for pair bond formation might be present in promiscuous vole species. Meadow voles do not in fact completely lack V1aRs in the ventral pallidum, and thus the potential to engage affiliative neural circuits is in place. One recent study found that a subpopulation of meadow voles formed selective affiliations under certain laboratory conditions, although this phenomenon has not been reported in other laboratories or in nature¹⁵. Our results show that overexpression of V1aRs to a degree beyond naturally occurring levels in meadow voles can shift every individual in the species to form pair bonds.

To determine whether similar proximate neural mechanisms underlie pair bond formation in V1aR-vp meadow voles and prairie voles, we next examined the role of dopaminergic neurotransmission. Before pairing with the same receptive female for 24 h, all animals were pretreated with 50 mg kg⁻¹ intraperitoneal eticlopride, a D2-receptor antagonist. This dose of eticlopride blocks partner preference formation in prairie voles¹⁶. Subsequent partner preference testing showed that V1aR-vp animals, like control animals, did not significantly prefer the partner over the stranger ($P > 0.05$, Student's *t*-test) (Fig. 4a). In addition, there were no significant group differences in the total time spent huddling with the partner ($P > 0.05$, Student's *t*-test). Analysis of the distribution of individuals within each group indicates that the distribution of control animals was unaffected by pretreatment with eticlopride, whereas the distribution of V1aR-vp animals was shifted back to the

normal continuum of control meadow voles, with no significant preference for the partner over that expected from random chance ($P > 0.05$, χ^2 analysis) (Fig. 4b). This shows that pair bond formation in the transgenic meadow voles, as in prairie voles, depends on dopaminergic neurotransmission.

Because monogamous-typical behaviours typically include biparental care in addition to pair bond formation and increased sociality, we proposed that V1aR overexpression in the ventral pallidum might result in increased paternal care by the males. There were no significant differences between the groups in 'classic' paternal behaviours, such as the latency to retrieval or the total time spent in physical contact (namely licking, grooming or huddling) with the pups. However, V1aR-vp animals did display increased social approach behaviours towards the pups, including a shorter latency to approach ($P < 0.05$, Student's *t*-test), more time near the pups ($P < 0.05$, Student's *t*-test), and less time self-grooming ($P < 0.005$, Student's *t*-test) than control animals, which is consistent with the finding of increased total social contact observed during the partner preference test. These results indicate that, although both behaviours are regulated by central V1aR, the underlying mechanisms of paternal care and pair bond formation probably use different neural circuits. In prairie voles, the administration of V1aR antagonist into the ventral pallidum selectively blocks partner preference⁴ but not paternal care; similarly, V1aR antagonist in the medial amygdala blocks paternal care but not partner preference (M.M.L. and L.J.Y., unpublished data). A similar behavioural dissociation between pair bond formation and paternal care has been reported in wild subpopulations of prairie voles: Illinois and Kansas prairie voles both form partner preferences, but only Illinois voles display high levels of paternal care¹⁷. Thus, the same V1aR gene that regulates the suite of monogamous-typical behaviours can have different effects on behaviours, depending on the specific neural circuits that are engaged.

What proximate mechanisms might underlie pair bond for-

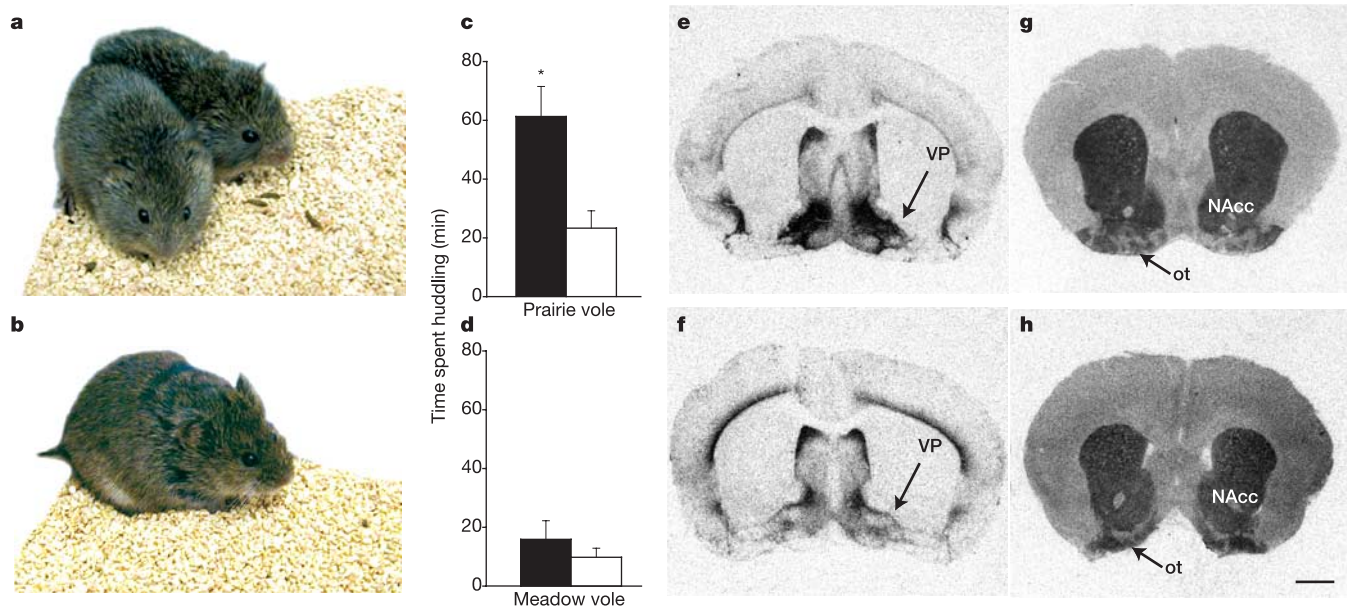


Figure 1 Comparison of brain neurochemistry and behaviour in prairie and meadow voles. **a, b**, Although prairie voles and meadow voles are similar in physical appearance, prairie voles are highly affiliative as depicted here in 'huddling' side by side (**a**), whereas meadow voles are solitary (**b**). **c, d**, Partner preference test. After mating and cohabitating with a female, a male prairie vole tended to spend significantly more time in contact with the partner (filled columns) than the stranger (open columns) ($P < 0.05$, Student's *t*-test) (**c**), whereas meadow voles do not form partner preferences and spent relatively little time

huddling with either female (**d**). Error bars, standard error. **e, f**, Autoradiograms of the ventral forebrain illustrating the typical prairie vole (**e**) and meadow vole (**f**) expression pattern of V1aR as shown by V1aR autoradiography. **g, h**, Despite considerable species differences in V1aR pattern, D2 receptor distribution was broadly similar in prairie voles (**g**) and meadow voles (**h**), as shown by D2 receptor binding in the nucleus accumbens (NAcc) and the olfactory tubercle (ot). Scale bar, 1 mm.

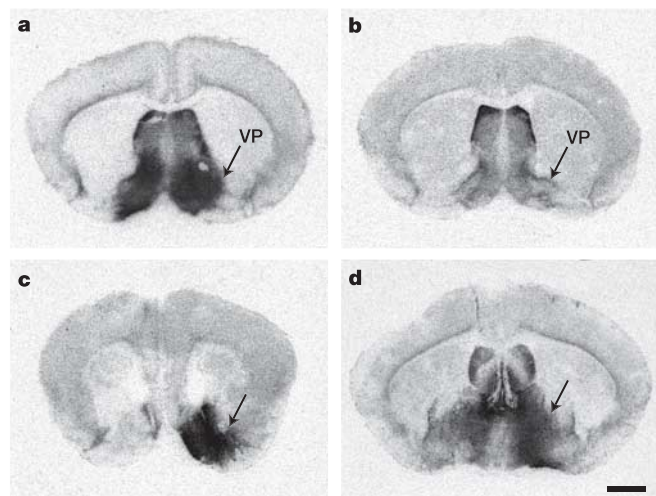


Figure 2 V1aR autoradiography at the level of the ventral pallidum. **a**, Meadow vole overexpressing the *V1aR* gene in the ventral pallidum by AAV-mediated gene transfer (V1aR-vp). **b**, Meadow vole infused with the AAV control vector expressing the *lacZ* gene into the ventral pallidum (Ctrl-vp). **c**, A stereotactic injection inadvertently placed too rostral to the ventral pallidum, in this case located just ventral to the nucleus accumbens. **d**, A stereotactic injection placed too caudal to the ventral pallidum, in this case just ventral to the fornix. Arrows depict ectopic AAV-mediated *V1aR* expression in **c** and **d**. Animals with AAV vector placement outside the ventral pallidum were placed in a second control group (Ctrl-other). Scale bar, 1 mm.

mation? V1aRs are also crucially involved in social memory formation. V1aR knockout mice display deficits in individual discrimination, whereas overexpression of *V1aR* in the rat brain by viral vector gene transfer increases the duration of social memory^{18,19}. We propose that, during pair bond formation, the concurrent activation of individual recognition and reward pathways results in convergent V1aR and D2 receptor activation in the ventral forebrain, leading to an association between the rewarding nature of sex and the olfactory signature of the partner, and thus the development of a conditioned partner preference. Similarly, the lack of V1aR in the ventral forebrain in promiscuous species might

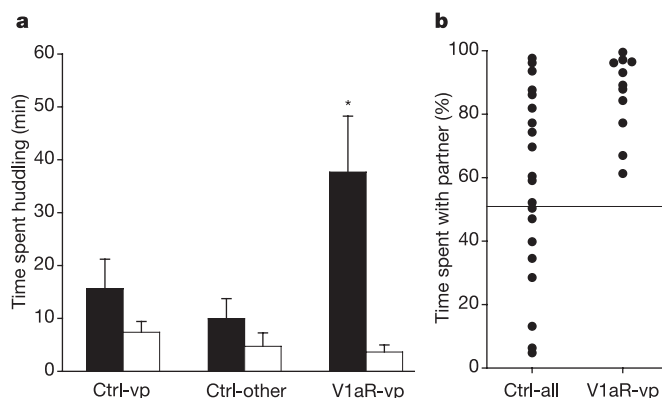


Figure 3 Partner preference test. **a**, V1aR-vp meadow voles spent significantly more time huddling with the partner (filled column) than the stranger (open column), whereas control animals (Ctrl-vp) and stereotactic misses (Ctrl-other) did not ($P < 0.01$, Student's *t*-test). Error bars, standard error. **b**, A plot of the percentage of time spent with the partner for each subject indicates a shift from randomly distributed preferences in the control groups to 100% of animals preferring the partner in the V1aR-vp group ($P < 0.001$, χ^2 analysis). The y axis was calculated as the time spent huddling with the partner divided by the total time spent huddling with the partner and stranger, multiplied by 100.

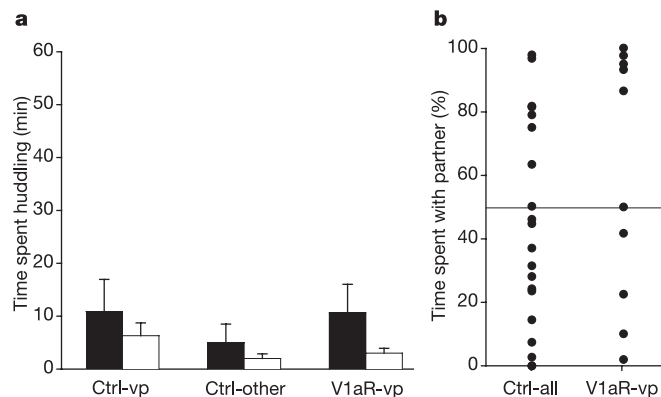


Figure 4 Partner preference test after eticlopride pretreatment. **a**, Eticlopride blocked partner preference in the V1aR-vp group. V1aR-vp, Ctrl-vp and Ctrl-other animals did not spend significantly more time with either the partner (filled columns) or the stranger (open columns) ($P > 0.05$, Student's *t*-test). Error bars, standard error. **b**, A plot of the percentage of time spent with the partner for each subject shows a similar distribution between control and V1aR-vp groups, with no significant difference from that expected by random chance ($P > 0.05$, χ^2 analysis).

lead to a lack of this selective association of reward with the partner's olfactory signature. In prairie voles, D1 receptors are upregulated after mating; the administration of a D1 receptor agonist inhibits pair bond formation, providing a potential mechanism for preventing the formation of subsequent pair bonds with novel individuals²⁰.

Although D2 receptor activation is required for pair bond formation in prairie voles, we did not observe any obvious species differences in D2 receptor patterns in the ventral forebrain (Fig. 1g, h). Thus, monogamous social organization might be the result of the insertion of the V1aR system into this ancient pre-existing reward circuit. Because our manipulation was performed on adult animals, this suggests that the pre-existing genetic complement and neural circuits underlying behaviour are fundamentally similar among vole species, and thus a simple alteration in the expression of a single gene can have a profound impact on behaviour.

The hypothesis that the vast majority of organismal complexity arises from differences in gene regulation, rather than from mutations in the coding regions themselves, has received much attention recently²¹. In prairie and pine voles, the 5' regulatory region of *V1aR* contains a microsatellite region that is virtually absent from montane and meadow voles⁹. This species-specific microsatellite polymorphism modulates gene expression in a cell-type-dependent manner²². Microsatellite DNA can rapidly contract and expand in length, providing a molecular mechanism for the generation of individual variability in brain V1aR patterns within the prairie vole species and between vole species^{23,24}. *V1aR* gene expression is exceptional in its phylogenetic plasticity in that no two species studied have identical V1aR distributions^{23,25}. Thus, the exceptional plasticity of V1aR patterns could result in an altered behavioural response to vasopressin, creating diversity in complex social behaviours and enabling adaptation to changing socioecological factors. It is noteworthy that similar genetic polymorphisms in the human *V1aR* promoter region exist, and one such allelic variant has been linked to autism²⁶, raising the possibility that human *V1aR* polymorphisms might also contribute to human variability in social behaviour.

We have shown that changes in the regional expression of a single gene can have a profound effect on the social behaviour of individuals within a species. Given a population of voles with variability in V1aR patterns and selection pressure for pair bond formation, one can see how monogamous social organization could

have evolved rapidly in voles. However, in a larger context, a single gene does not act alone in the control of complex social behaviour; it must ultimately be placed within pre-existing biological pathways that then interact with socioecological factors, developmental pathways and stochastic events in the lives of organisms. □

Methods

Animals and treatment

Meadow and prairie voles were laboratory-reared animals derived from field-caught specimens. Subjects were 2–6-month-old sexually naive male voles, 40–70 g, housed with one or two same-sex littermates in a 14:10 light:dark cycle and were provided with rabbit chow and water *ad libitum*. Experiments were performed at Emory University in compliance with the rules and oversight of the Emory Institutional Use and Care of Animals Committee.

The V1aR-vp group of male meadow voles ($n = 16$) received bilateral infusions into the ventral pallidum of the adeno-associated viral (AAV) vector containing the prairie vole V1aR gene and under the control of a neuron-specific enolase promoter as described previously²⁷. The Ctrl-vp group (control; $n = 15$) received bilateral infusions into the ventral forebrain of the control AAV vector expressing the lacZ gene under a cytomegalovirus promoter. The prairie vole V1aR and *Escherichia coli* lacZ recombinant viral vectors were cloned and packaged as described previously²⁷. AAV infusions were performed under isoflurane anaesthesia in a Kopf stereotax fitted with an Ultra Micro Pump II (World Precision Instruments) and 26-gauge Hamilton syringe. Stereotactic coordinates were determined beforehand with dye injections, and varied depending on the weight of the animal (rostral, 1.5 mm; bilateral, 0.9 mm; ventral, 5.8 mm relative to Bregma). Once the syringe had been lowered to the injection site, the AAV vector was infused at a rate of 3–5 nl s⁻¹ for a total of 1 µl per side, at a viral titre of 10⁸ infectious units µl⁻¹, as described previously²⁷. Animals woke from the anaesthesia within 5 min of surgery and were allowed to recover for 2 weeks before behavioural testing.

In advance of cohabitation before the third partner preference test, all animals were pretreated with an intraperitoneal injection of 50 mg kg⁻¹ eticlopride, a selective D2-receptor antagonist, dissolved in 0.1 ml lactated Ringer's solution (Sigma). All animals were observed to mate during the 24-h cohabitation.

The animals used in Fig. 1 were sexually naive male prairie voles ($n = 19$) and sexually naive male meadow voles ($n = 10$) who were cohabitated with a sexually receptive conspecific female for 24 h and then immediately tested for partner preference.

Behavioural testing

Partner preference tests were performed immediately after a 24-h cohabitation with a behaviourally receptive conspecific female. Behaviour during the first hour of cohabitation was recorded for the number of mating bouts; V1aR-vp and Ctrl-vp animals did not differ significantly in the total number of mating bouts in any of the cohabitation periods. During the partner preference test, the experimental male was placed in the centre, neutral chamber of a three-chambered apparatus in which the partner female was tethered in one chamber and a novel stranger female was tethered in the second. The experimental animal was free to move throughout the chambers through Plexiglas (Perspex) connecting tubes. The time spent in each cage and the time spent huddling with each female were recorded by an experimenter blind to the treatment groups.

To show that the animals could be tested multiple times in the partner preference test, we separated the animals for 2 weeks, then paired them with the same receptive female for 24 h, and subsequently tested them again in the partner preference test. The results from the first partner preference test were replicated; thus, there was no significant effect of retesting the animals. In total, three sets of partner preference tests were conducted. All sets of partner preference tests were conducted 2 weeks apart. The third partner preference test consisted of pretreatment with eticlopride before the 24 h cohabitation with a receptive female.

For the paternal care test, because of limitations in the availability of meadow vole pups within the proper age window, only six subjects in each experimental group were chosen at random and tested. In brief, two pups between 2 and 5 days of age were placed in the opposite corner of the subject's home cage for 10 min. An experimenter, blind to the treatment group of each animal, recorded behaviours for latency to approach, latency to retrieval, time spent near the pups (within 5 cm), time spent in contact with the pups (licking, grooming, crouching), and time spent self-grooming.

Receptor autoradiography

V1aR expression in the ventral forebrain was revealed by receptor autoradiography with ¹²⁵I-linear-AVP antagonist (PerkinElmer/NEN) as described previously²⁸.

D2 receptor expression in the forebrain was detected by receptor autoradiography with ¹²⁵I-iodospiperone (PerkinElmer/NEN). In brief, brains were removed, snap-frozen on dry ice, and sliced on a cryostat at 20 µm on Superfrost Plus slides (Fisher Scientific). Slides were washed twice for 10 min each in 50 mM Tris-Cl buffer pH 7.4, then incubated for 1 h in 50 pM radiolabelled ligand dissolved in 5.7 mM ascorbic acid and Tris-ions buffer (0.7% NaCl, 0.04% KCl, 0.02% CaCl₂, 0.01% MgCl₂ in Tris buffer, pH 7.4). Slides were then washed four times (5 min each) in Tris-ions buffer followed by a final stirring, 30 min wash, and dipped in distilled water. After being dried with a stream of cool air, slides were apposed to Kodak Biomax MR film for 24 h.

Received 29 January; accepted 2 April 2004; doi:10.1038/nature02539.

- Gruder-Adams, S. & Getz, L. L. Comparison of the mating system and paternal behavior in *Microtus ochrogaster* and *M. pennsylvanicus*. *J. Mamm.* **66**, 165–167 (1985).
- Aragona, B. J., Liu, Y., Curtis, J. T., Stephan, F. K. & Wang, Z. A critical role for nucleus

accumbens dopamine in partner-preference formation in male prairie voles. *J. Neurosci.* **23**, 3483–3490 (2003).

- Gingrich, B., Liu, Y., Cascio, C., Wang, Z. & Insel, T. R. Dopamine D2 receptors in the nucleus accumbens are important for social attachment in female prairie voles (*Microtus ochrogaster*). *Behav. Neurosci.* **114**, 173–183 (2000).
- Lim, M. M. & Young, L. J. Vasopressin-dependent neural circuits underlying pair bonding in the monogamous prairie vole. *Neuroscience* **125**, 35–45 (2004).
- Insel, T. R., Wang, Z. X. & Ferris, C. F. Patterns of brain vasopressin receptor distribution associated with social organization in microtine rodents. *J. Neurosci.* **14**, 5381–5392 (1994).
- Kleiman, D. G. Monogamy in mammals. *Q. Rev. Biol.* **52**, 39–69 (1977).
- Winslow, J. T., Hastings, N., Carter, C. S., Harbaugh, C. R. & Insel, T. R. A role for central vasopressin in pair bonding in monogamous prairie voles. *Nature* **365**, 545–548 (1993).
- Wang, Z., Ferris, C. F. & De Vries, G. J. Role of septal vasopressin innervation in paternal behavior in prairie voles (*Microtus ochrogaster*). *Proc. Natl Acad. Sci. USA* **91**, 400–404 (1994).
- Young, L. J., Nilsen, R., Waymire, K. G., MacGregor, G. R. & Insel, T. R. Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. *Nature* **400**, 766–768 (1999).
- Lim, M. M., Murphy, A. Z. & Young, L. J. Ventral striatopallidal oxytocin and vasopressin V1a receptors in the monogamous prairie vole (*Microtus ochrogaster*). *J. Comp. Neurol.* **468**, 555–570 (2004).
- Young, L. J., Toloczko, D. & Insel, T. R. Localization of vasopressin (V1a) receptor binding and mRNA in the rhesus monkey brain. *J. Neuroendocrinol.* **11**, 291–297 (1999).
- Young, L. J. Frank A. Beach Award. Oxytocin and vasopressin receptors and species-typical social behaviors. *Horm. Behav.* **36**, 212–221 (1999).
- Wang, Z. *et al.* Vasopressin in the forebrain of common marmosets (*Callithrix jacchus*): studies with *in situ* hybridization, immunocytochemistry and receptor autoradiography. *Brain Res.* **768**, 147–156 (1997).
- Bester-Meredith, J. K., Young, L. J. & Marler, C. A. Species differences in paternal behavior and aggression in peromyscus and their associations with vasopressin immunoreactivity and receptors. *Horm. Behav.* **36**, 25–38 (1999).
- Parker, K. J. & Lee, T. M. Development of selective partner preferences in captive male and female meadow voles, *Microtus pennsylvanicus*. *Anim. Behav.* **61**, 1217–1226 (2001).
- Wang, Z. *et al.* Dopamine D2 receptor-mediated regulation of partner preferences in female prairie voles (*Microtus ochrogaster*): a mechanism for pair bonding? *Behav. Neurosci.* **113**, 602–611 (1999).
- Roberts, R. L., Williams, J. R., Wang, A. K. & Carter, C. S. Cooperative breeding and monogamy in prairie voles: influence of the sire and geographical variation. *Anim. Behav.* **55**, 1131–1140 (1998).
- Bielsky, I. F., Bao-Hu, S., Szegda, K. L., Westphal, H. & Young, L. J. Profound impairment in social recognition and reduction in anxiety-like behavior in vasopressin V1a receptor knockout mice. *Neuropsychopharmacology* **29**, 483–493 (2004).
- Landgraf, R. *et al.* Viral vector-mediated gene transfer of the vole V1a vasopressin receptor in the rat septum: improved social discrimination and active social behaviour. *Eur. J. Neurosci.* **18**, 403–411 (2003).
- Aragona, B. J. *et al.* Opposite modulation of social attachment by D1- and D2-type dopamine receptor activation in nucleus accumbens shell. *Horm. Behav.* **44**, 37 (2003).
- Robinson, G. E. & Ben-Shahar, Y. Social behavior and comparative genomics: new genes or new gene regulation? *Genes Brain Behav.* **1**, 197–203 (2002).
- Hammock, E. A. D. & Young, L. J. Functional microsatellite polymorphism associated with divergent social structure in vole species. *Mol. Biol. Evol.* (in the press).
- Phelps, S. M. & Young, L. J. Extraordinary diversity in vasopressin (V1a) receptor distributions among wild prairie voles (*Microtus ochrogaster*): patterns of variation and covariation. *J. Comp. Neurol.* **466**, 564–576 (2003).
- Hammock, E. A. D. & Young, L. J. Variation in the vasopressin V1a receptor promoter and expression: implications for inter- and intraspecific variation in social behaviour. *Eur. J. Neurosci.* **16**, 399–402 (2002).
- Barberis, C. & Tribollet, E. Vasopressin and oxytocin receptors in the central nervous system. *Crit. Rev. Neurobiol.* **10**, 119–154 (1996).
- Kim, S. J. *et al.* Transmission disequilibrium testing of arginine vasopressin receptor 1A (AVPR1A) polymorphisms in autism. *Mol. Psychiatry* **7**, 503–507 (2002).
- Pitkow, L. J. *et al.* Facilitation of affiliation and pair-bond formation by vasopressin receptor gene transfer into the ventral forebrain of a monogamous vole. *J. Neurosci.* **21**, 7392–7396 (2001).
- Young, L. J., Winslow, J. T., Nilsen, R. & Insel, T. R. Species differences in V1a receptor gene expression in monogamous and nonmonogamous voles: behavioral consequences. *Behav. Neurosci.* **111**, 599–605 (1997).

Acknowledgements We thank C. A. Sharer for contributing to the initial viral vector work, L. Miller for assistance with meadow vole breeding, and A. Z. Murphy and E. A. D. Hammock for their critical readings of the manuscript. Funding for this work was provided by grants from the NIH to M.M.L., Z.X.W., L.J.Y. and the Yerkes Center, and by a NSF STC grant to the Center for Behavioral Neuroscience.

Authors' contributions M.M.L. performed the experiments and wrote the manuscript. Z.X.W. provided the animals and behavioural testing equipment for the pilot studies, the D2 autoradiography protocol and scientific input. D.E.O. assisted with the paternal behaviour testing. X.H.R. and E.E.T. generated the viral vectors. L.J.Y. conceived the idea and co-wrote the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.J.Y. (lyoun03@emory.edu).

Local-feature assembling in visual pattern recognition and generalization in honeybees

Silke Stach^{1,2}, Julie Benard¹ & Martin Giurfa¹

¹Research Centre for Animal Cognition, CNRS—Université Paul-Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 4, France

²Neurobiologie, Institut für Biologie, Freie Universität Berlin, Königin-Luise-Str. 28/30, 14195 Berlin, Germany

Generalization is a cognitive ability that allows similar stimuli along a given dimension to be treated as equivalents^{1–3}. Insects exhibit high levels of visual generalization^{4–6}. Honeybees trained to recognize complex visual stimuli on the basis of a single feature generalize their choice to novel stimuli sharing that common feature with the trained stimuli⁷. The demonstration of this kind of performance has been limited to the use of a single visual feature, and the possibility that bees link different features in learning a visual pattern has been denied^{8,9}. Here we show that honeybees trained with a series of complex patterns sharing a common layout comprising four edge orientations remember these orientations simultaneously in their appropriate positions, and generalize their response to novel stimuli that preserve the trained layout. Honeybees also generalize their response to patterns with fewer correct orientations, depending on their match with the trained layout. Stimulation of the achromatic L-photoreceptor input is necessary for this task. The mini-brain of the honeybee can thus extract regularities in its environment and establish correspondences among correlated features. It can thus generate a large set of object descriptions from a finite set of elements.

Honeybees are well suited for studies of visual recognition as they can be easily trained to fly towards a given stimulus to obtain a reward of sucrose solution on it¹⁰. Honeybees trained to choose patterns on the basis of a single feature such as size¹¹, bilateral¹² or radial symmetry¹³, concentric pattern organization¹³, edge orientation¹⁴, or pattern disruption¹⁵ generalize their response to previously unseen patterns that share the trained feature. Single-feature learning in honeybees can also have a spatial component, as the position and orientation of an edge in specific regions of the visual field can be learned if they remain constant during training¹⁶. Despite the high level of generalization underlying these performances, recognition based on a single feature is a rather limited achievement when compared with visual recognition exhibited by humans¹⁷ in which subjects can learn and discriminate objects on the basis of their relative arrangement of multiple parts or features. Here we investigated whether honeybees can learn and recognize patterns on the basis of a layout of linked features and whether they generalize their choice to novel stimuli sharing the same basic layout of features as the trained patterns. Furthermore, we determined the physiological pathway involved in this task.

Individually marked honeybees, *Apis mellifera*, were trained with a series of six randomly changing pairs of patterns presented vertically on the back walls of a naturally illuminated Y-shaped maze. Only one honeybee was present in the maze at a time. The stimulus defined by the experimenter as positive was rewarded with sucrose solution while that defined as negative was non-rewarded. Under such a differential conditioning, honeybees learn the patterns as a whole, not only their local cues¹⁸. Patterns were disks divided into four quadrants, each exhibiting a different orientation. Depending on the experiment, the training patterns were either black and white on a white background (Fig. 1a) or coloured on a grey background (not shown). Bar width and grating period could

be well resolved by the honeybees within the maze¹⁹. Patterns labelled 'A' shared a common layout defined by the four different orientations. Patterns labelled 'B' also shared a common orientation layout that was perpendicular to that of the A patterns (Fig. 1a).

In a first experiment, honeybees were trained during 42 trials (that is, 42 foraging bouts) with a random succession of the six pairs of black and white A and B patterns (Fig. 1a). For one group of honeybees, the A patterns were rewarded and the B patterns were not, and vice versa for the other group. Both groups learned equally well to choose the appropriate stimulus within each pair. Pooled acquisition increased significantly and reached a proportion of 0.67 correct choices at the end of training (Fig. 2a; $F = 5.2$, degrees of freedom (d.f.) = 6,96, $P < 0.001$). To test whether honeybees were capable of extracting the layout common to the rewarded patterns, they were presented with novel test stimuli that were all non-rewarded and had not been seen by the honeybees beforehand. In each quadrant, orientation cues were reduced to a single oriented bar (Fig. 1b). One of the test stimuli always preserved the previously rewarded layout (positive layout S+) whereas the other differed in a single orientation. To this end, the orientation of one of the four bars was rotated by 90° with respect to the positive layout (Fig. 1b). Each of the four possible alternatives was contrasted to the simplified positive layout in a separate test. These tests allow deciding whether or not honeybees can extract and remember simultaneously the four orientations in their corresponding quadrants, in which case they should always prefer the complete positive layout S+ to any stimulus in which one of the four bars was rotated. Honeybees of both groups preferred the test patterns with the complete positive layout S+ to any of the four modified alternatives, regardless of the quadrant in which the bar was rotated (Fig. 2b; $P < 0.001$ in all four cases). As we originally counted the contacts with the test stimuli, the honeybees could have decided when they were very close to the patterns and might have therefore relied only on local cues and not on the global layout of four edge orientations. We tested this possibility by repeating the previous experiment and placing a transparent barrier with a small, central entry hole cut into it at the entrance of each arm of the maze¹⁵. Before choosing, honeybees had now to fixate each pattern at a distance of 13 cm, which corresponds to an unambiguous measure of stimulus retinal size of 37° at the choice point. In the training as in the tests, we recorded only the first entry into one or other arm of the maze. The results of this control experiment did not differ from those of the first experiment (not shown). Thus, the honeybees were able to recognize patterns on the basis of a common layout made up of four different edge orientations, and used such a simplified representation to respond to novel patterns. Within this simplified representation, all four quadrants had the same importance.

In a second experiment we trained the honeybees as in the first experiment to promote learning of a common layout of edge orientations. Acquisition again increased significantly and reached a proportion of 0.71 correct choices at the end of training (Fig. 2c; $F = 5.4$, d.f. = 6,72, $P < 0.001$). Such performance did not differ from that observed in the first experiment ($F = 0.42$, d.f. = 1,28, not significant (NS); compare Fig. 2a and c). We then tested pattern-layout generalization by confronting honeybees with three pairs of novel test patterns differing in their appearances but not in their orientation layouts (Fig. 1c). In all cases, honeybees preferred the novel pattern corresponding to the previously rewarded layout (Fig. 2d; $P < 0.001$ in all cases). They were thus able to respond appropriately to different kinds of unknown patterns on the basis of a common layout of edge orientations.

In a third experiment we investigated whether honeybees trained to recognize patterns on the basis of the four-edge layout also respond to patterns that exhibit an incomplete layout. If honeybees choose patterns on the basis of a best match between the memorized and the perceived layout, they should prefer the incomplete layout that most closely resembles the trained, complete layout. Honeybees

were trained as in the first experiment and then presented with novel test stimuli that were all non-rewarded. One of the test stimuli always preserved the previously non-rewarded layout (Fig. 1d; negative layout S−), which displayed 0 correct edges. In four separate tests, this test stimulus was presented against one of four different alternatives (UL, UR, LL and LR; see Fig. 1d for definitions), which displayed three correct edges each. Acquisition increased significantly and reached a proportion of 0.63 correct choices at the end of training (Fig. 2e; $F = 3.98$, d.f. = 6,36, $P < 0.005$). In the tests, honeybees always preferred the patterns with three correct edges to those with 0 correct edges (Fig. 2f; $P < 0.05$ in all four cases). Thus, honeybees generalize their

response to stimuli that do not exhibit the complete memorized layout of orientations but that allow a best match with it.

We next studied the involvement of different spectral inputs in this visual task. Three kinds of photoreceptor exist in the honeybee retina: S, M and L, with maximum sensitivity in the ultraviolet (S or ultraviolet receptor; $\lambda_{\max} = 344$ nm), blue (M or blue receptor; $\lambda_{\max} = 436$ nm) and green regions of the spectrum (L or green receptor; $\lambda_{\max} = 544$ nm), respectively²⁰. Global orientation of a pattern is processed through the achromatic channel of the L receptor alone (green receptor type)²¹. In a fourth experiment, we trained honeybees as in the first experiment but the training stimuli were coloured on a grey background. Three colours were chosen to

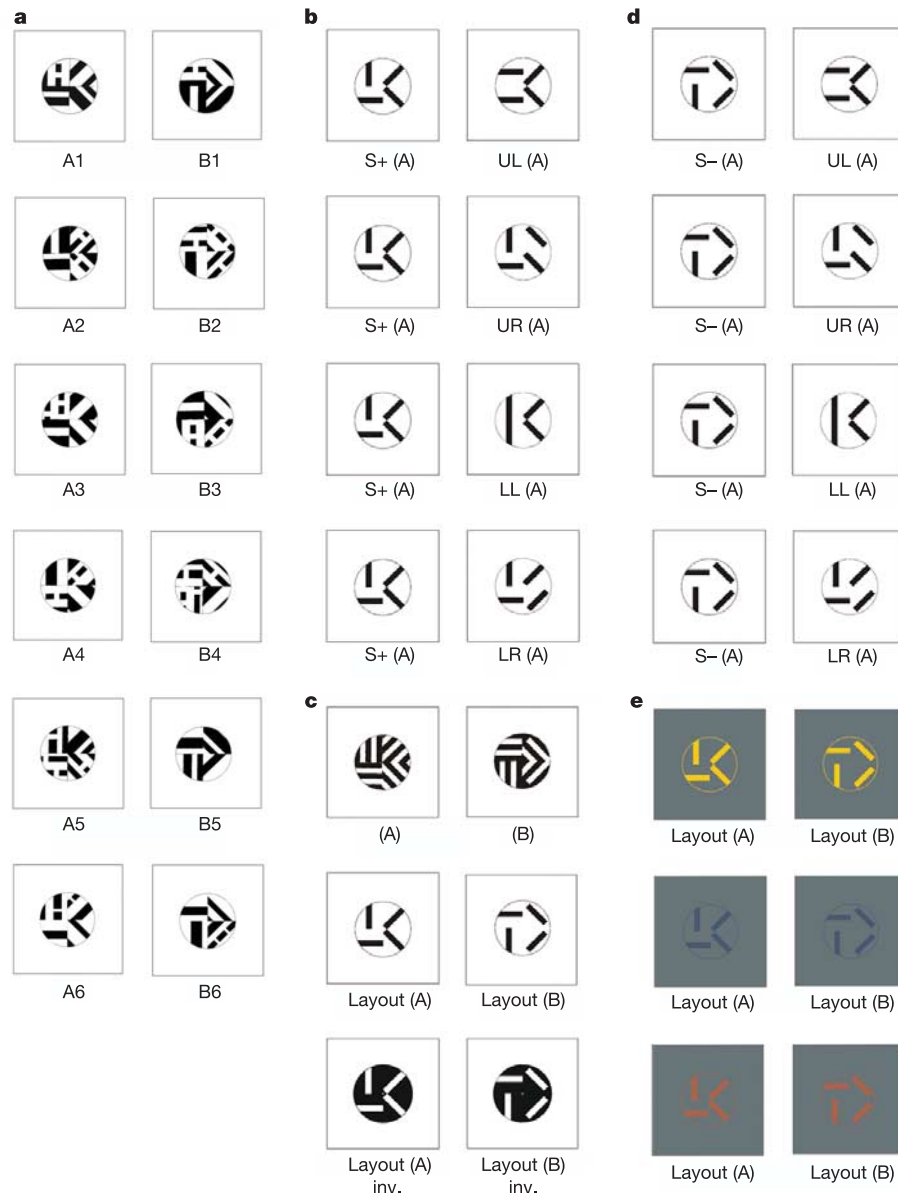


Figure 1 Training and test stimuli. **a**, Black and white training stimuli in the first to third experiments. A patterns (A1 to A6) differed from each other but shared a common layout defined by the spatial arrangement of orientations in the four quadrants. B patterns (B1 to B6) shared a common layout perpendicular to that of A patterns. In the fourth experiment, the training patterns were coloured against a grey background (not shown). **b**, Test stimuli in the first experiment. The four test pairs shown correspond to the honeybees trained with A patterns. S+, simplified layout of the rewarded training patterns; UL, upper-left bar rotated; UR, upper-right bar rotated; LL, lower-left bar rotated; LR, lower-right bar rotated. Equivalent tests were performed with the honeybees trained

with B patterns (not shown). **c**, Test stimuli in the second experiment. Layout, simplified layout of training patterns; layout inv., layout with black and white areas inverted. **d**, Test stimuli in the third experiment. The four test pairs shown correspond to the group of honeybees trained with A patterns. S−, simplified layout of the non-rewarded training patterns (B patterns in this case). UL, UR, LL and LR correspond to variations of S+ in which a single bar was rotated (see above). Equivalent tests were performed with honeybees trained with B patterns (not shown). **e**, Coloured test patterns used in the fourth experiment.

provide different receptor-specific contrasts to the background: yellow patterns in which contrast to the L-receptor type was dominant; blue ones in which contrasts to the S- and the M-receptor type were not available but presenting a small negative L contrast; and brown ones in which contrast to the L-receptor type was practically absent (see Table 1). Honeybees could only learn to choose the appropriate stimulus within each of the six training pairs in the case of the yellow stimuli, which provided dominant L contrast. Performance increased significantly during training and reached a proportion of 0.84 correct choices in the last block (Fig. 3a; $F = 10.7$, d.f. = 6,72, $P < 0.001$). The task could not be learned in the case of the blue and brown stimuli (Fig. 3a; blue patterns: $F = 1.6$, d.f. = 6,66, NS; brown patterns: $F = 1.8$, d.f. = 6,72, NS) although a positive tendency was visible at the end of training. In the case of yellow stimuli, acquisition was significantly better than that with the achromatic patterns ($F = 25.23$, d.f. = 1,41, $P < 0.001$; compare with Fig. 2a), thus showing that chromatic cues improve performance in this task. Tests with simplified yellow, blue and brown patterns (Fig. 1d) showed that honeybees preferred the

correct simplified layout ($P < 0.005$, Fig. 3b) only in the case of the yellow stimuli in which L contrast was dominant. When the test patterns were blue or brown and presented low or no L contrast, honeybees chose randomly between them or slightly preferred the wrong alternative, respectively. Thus, absence of L contrast affected learning of patterns in terms of a layout of four edge orientations.

Our results show that honeybees can recognize visual patterns on the basis of a global layout made from four different orientations, common to a series of different patterns. Honeybees built a generic stimulus representation in which the four orientations coexisted in a specific spatial layout, which allowed appropriate responses to novel stimuli exhibiting or not this layout. Generalization depended on a best match between the memorized and the perceived layout and thus on the number of coincident edges between layouts. As the honeybees transferred their choice to unknown stimuli, visual recognition was not driven by retinotopic matching with a template in a strict sense²². Randomizing the patterns during training precludes the building of specific templates and generalization should be reduced if honeybees memorize patterns on a pixel basis. Honeybees might therefore extract and link different visual features together in order to build simplified representations for objects such as flowers, plants, landmarks and other objects in their environment. The small brain of the honeybee, with its 950,000 neurons, could be therefore capable of generating large sets of object descriptions from finite sets of elements.

Suppression or reduction of L-receptor contrast impaired the performance of the honeybees. This result is consistent with the importance of L contrast for edge detection²¹. Orientation detectors with small visual fields are required for the implementation of an orientation layout like the one demonstrated by our experiments. Such detectors, with receptive fields from 2° to 20°, have been found in the insect brain^{23,24}. It is conceivable that their simultaneous activation provides the basis for pattern representation in terms of a specific set of orientations. The orientations available in the four

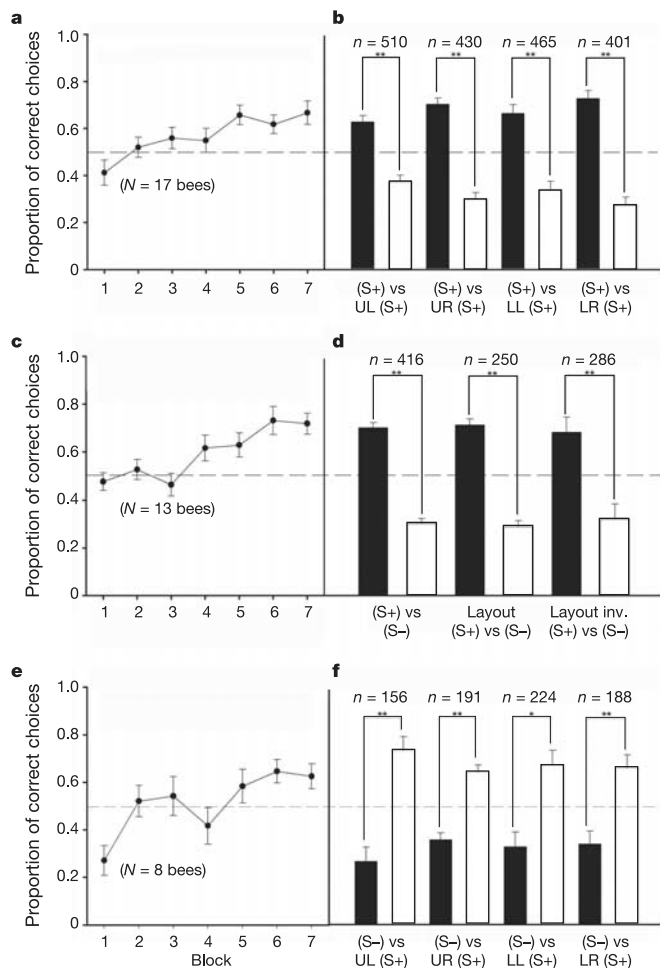


Figure 2 First (a, b), second (c, d) and third (e, f) experiments. S+ and S- indicate layouts of the training patterns previously rewarded and non-rewarded, respectively. **a, c, e**, Acquisition in the first, second and third experiment, respectively. The graphs show the proportion of correct choices in seven blocks of six consecutive visits. **b, d, f**, Proportion of correct choices in the tests of the first, second and third experiment, respectively. In the first and second experiment, honeybees preferred the simplified layout of the training patterns previously rewarded (S+); in the third experiment, they preferred the incomplete layouts with three correct edges to the simplified layout previously non-rewarded (S-) with no correct edges. Error bars indicate standard error of the mean. *N* indicates the number of honeybees, whereas *n* indicates the number of touches. Asterisk, $P < 0.05$; double asterisk, $P < 0.005$.

Table 1 Relative receptor quantum catches for the coloured stimuli used			
Stimuli	Relative receptor quantum catches		
	S	M	L
Background	1.0	1.0	1.0
Yellow	1.1	1.4	3.0
Brown	0.6	0.4	0.9
Blue	1.0	1.0	0.7

Relative receptor quantum catches are normalized with respect to the background. Bold values indicate coincidence in receptor quantum catches; that is, lack or reduction of achromatic contrast. Yellow, brown and blue refer to human colour perceptions.

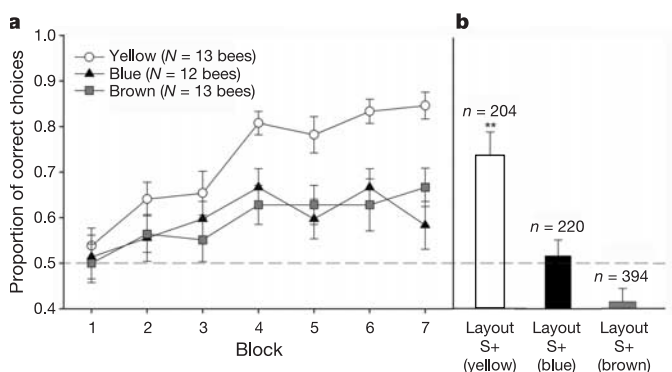


Figure 3 Fourth experiment. **a**, Acquisition during training with six pairs of chromatic patterns (yellow, blue or brown) identical to those of Fig. 1a. The figure shows the proportion of correct choices in seven blocks of six consecutive visits. **b**, Proportion of correct choices in the tests (see Fig. 1d). Error bars indicate standard error of the mean. *N* indicates number of honeybees, whereas *n* indicates number of touches. Double asterisk, $P < 0.005$.

quadrants might excite differently oriented edge detectors and this information might be combined to define the correct (and incorrect) layout. Honeybees will respond appropriately to novel stimuli that excite the same or partial arrangements of detectors.

The belief that insects have limited visual recognition capabilities is questioned by our results. Topological invariants of visual targets are relevant, global properties in insect vision²⁵. Furthermore, our results show that integration of local features is a strategy that is also available to the insect's visual system. Despite the apparent simplicity of their nervous systems, insects can access visual recognition strategies that yield surprising similarities to those used by vertebrates^{26–28}. □

Methods

Individually marked honeybees were trained to collect 50% sucrose solution in a Y-shaped maze. The maze was covered by an ultraviolet-transparent Plexiglas ceiling and shielded from direct sunlight. The entrance led to a decision chamber, where the honeybee could choose between two arms. Each arm was 40 × 20 × 20 cm. During training and tests only one honeybee was present in the apparatus.

Stimuli

Disks of 10 cm in diameter were presented vertically at a distance of 15 cm from the entrance hole, thus subtending a visual angle of 37°. The black-and-white patterns were produced with a high-resolution laser printer on copy paper of constant quality. The coloured patterns were printed with a high-resolution ink-jet printer on the same paper. The patterns were divided into four quadrants, each presenting a grating of a different orientation (0°, 45°, 90° and 135° with respect to the vertical). The width of stripes varied but it was always at least 1 cm, which corresponds to a visual angle of 4° as seen from the maze entrance. Stripes were always well resolvable for the honeybees' eyes¹⁹. The reward was provided in the centre of the patterns.

Colour parameters

The reflectance spectra of the coloured stimuli were measured by means of a spectrophotometer (Ocean Optics SD2000 with a DT1000 mini light source (200–1,100 nm) and R400-7 UV/VIS optical fibre). The receptor quantum catches, Q_i , were calculated as follows:

$$Q_i = \int_{300}^{700} I(\lambda) S_i(\lambda) R(\lambda) d(\lambda)$$

where subscript i is receptor S, M or L; λ is the wavelength; $I(\lambda)$ is the illumination spectrum (D65 standard function); $S_i(\lambda)$ is the spectral sensitivity function of receptor i ²⁰; and R is the measured reflectance spectrum of the coloured stimulus. For each colour/background combination the receptor-specific contrasts q_i (see Table 1) were calculated as $q_i = Q_i^t/Q_i^b$ where Q_i^t and Q_i^b are the quantum catches of receptor i for the target and the background colours.

Procedure

During training the side of the rewarded pattern (left or right) was interchanged to avoid positional learning. Each of the six pairs of training patterns was presented seven times in a randomized sequence, thus resulting in 42 training trials. Previous controls showed that this amount of training is sufficient to ensure learning in this task. Acquisition curves were established by dividing the 42 trials into 7 blocks of 6 trials each. Choice proportions were calculated from the number of choices for the positive stimuli (correct choices) per block. After the training, transfer tests with different non-rewarded patterns were performed. Contacts with the surface of the patterns were counted for 2 min. The choice proportion for each of the two test patterns was calculated. Each test was done twice, interchanging the sides of the patterns to control for side preferences. Refreshing trials were intermingled among the tests.

One-way analysis of variance (ANOVA) was performed to detect differences in performance between honeybees rewarded on A patterns and those rewarded on B patterns. In all cases, no significant differences were found, such that data were pooled for all subsequent treatments. ANOVAs for repeated measures were performed to detect whether acquisition significantly improved during training. To detect differences in test performances, a single score was derived from each 2-min test and was compared to a theoretical proportion of 50% by means of a one-sided Mann–Whitney U -test. When only the first choice of the honeybees was recorded in the tests, a binomial test was used. The α -level was set to 0.05 for all analyses.

Received 25 November 2003; accepted 21 April 2004; doi:10.1038/nature02594.

- Spence, K. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.* **44**, 430–444 (1937).
- Shepard, R. N. Toward a universal law of generalization for psychological science. *Science* **237**, 1317–1323 (1987).
- Ghirlanda, S. & Enquist, M. A century of generalization. *Anim. Behav.* **66**, 15–36 (2003).
- Srinivasan, M. V. Pattern recognition in the honeybee: recent progress. *J. Insect Physiol.* **40**, 183–194 (1994).
- Ronacher, B. How do bees learn and recognize visual patterns? *Biol. Cybern.* **79**, 477–485 (1998).
- Giurfa, M. & Menzel, R. Insect visual perception: complex ability of a simple nervous system. *Curr. Opin. Neurobiol.* **7**, 505–513 (1997).
- Giurfa, M. Cognitive neuroethology: dissecting non-elemental learning in a honeybee brain. *Curr. Opin. Neurobiol.* **13**, 726–735 (2003).

- Horridge, G. A. Vision of the honeybee *Apis mellifera* for patterns with two pairs of equal orthogonal bars. *J. Insect Physiol.* **42**, 131–138 (1996).
- Horridge, G. A. Vision of the honeybee *Apis mellifera* for patterns with one pair of equal orthogonal bars. *J. Insect Physiol.* **43**, 741–748 (1997).
- von Frisch, K. *The Dance Language and Orientation of Bees* 566 (Belknap Press of Harvard Univ. Press, Cambridge, Massachusetts, 1967).
- Ronacher, B. Pattern recognition in honeybees: multidimensional scaling reveals a city-block metric. *Vision Res.* **32**, 1837–1843 (1992).
- Giurfa, M., Eichmann, B. & Menzel, R. Symmetry perception in an insect. *Nature* **382**, 458–461 (1996).
- Horridge, G. A. & Zhang, S. W. Pattern vision in honeybees (*Apis mellifera*): Flower-like patterns with no predominant orientation. *J. Insect Physiol.* **41**, 681–688 (1995).
- van Hateren, J. H., Srinivasan, M. V. & Wait, P. B. Pattern recognition in bees: orientation discrimination. *J. Comp. Physiol.* **A167**, 649–654 (1990).
- Horridge, G. A. Pattern discrimination by the honeybee: disruption as a cue. *J. Comp. Physiol.* **A 181**, 267–277 (1997).
- Horridge, G. A. The effect of complexity on the discrimination of oriented bars by the honeybee (*Apis mellifera*). *J. Comp. Physiol.* **A 189**, 703–714 (2003).
- Maurer, D., Le Grand, R. & Mondloch, C. J. The many faces of configural processing. *Trends Cogn. Sci.* **6**, 255–260 (2002).
- Giurfa, M. et al. Pattern learning by honeybees: conditioning procedure and recognition strategy. *Anim. Behav.* **57**, 315–324 (1999).
- Srinivasan, M. V. & Lehrer, M. Spatial acuity of honeybee vision and its spectral properties. *J. Comp. Physiol.* **A 162**, 159–172 (1988).
- Menzel, R. & Backhaus, W. in *Vision and Visual Dysfunction. The Perception of Colour* (ed. Gouras, P.) 262–288 (MacMillan, London, 1991).
- Giger, A. D. & Srinivasan, M. V. Pattern recognition in honeybees: chromatic properties of orientation analyses. *J. Comp. Physiol.* **A 178**, 763–769 (1996).
- Dill, M., Wolf, R. & Heisenberg, M. Visual pattern recognition in *Drosophila* involves retinotopic matching. *Nature* **365**, 751–753 (1993).
- Osorio, D. Temporal and spectral properties of sustaining cells in the medulla of the locust. *J. Comp. Physiol.* **161**, 441–448 (1987).
- James, A. C. & Osorio, D. Characterisation of columnar neurons and visual signal processing in the medulla of the locust optic lobe by system identification techniques. *J. Comp. Physiol.* **A 178**, 183–199 (1996).
- Chen, L., Zhang, S. & Srinivasan, M. V. Global perception in small brains: topological pattern recognition in honeybees. *Proc. Natl Acad. Sci. USA* **100**, 6884–6889 (2003).
- Biedermann, I. Recognition-by-components: A theory of human image understanding. *Psychol. Rev.* **94**, 115–147 (1987).
- Werner, C. W. & Rehämper, G. Discrimination of multidimensional geometrical figures by chickens: categorization and pattern learning. *Anim. Cogn.* **2**, 27–40 (1999).
- Treisman, A. & Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980).

Acknowledgements We thank J. Delius, M. Fabre-Thorpe, L. Huber, R. Menzel and S. Thorpe for valuable comments and corrections on previous versions of the manuscript, and N. Hempel de Ibarra for help with colour measurements. This work was supported by the German (DFG) and the French (CNRS) Research Councils, the University Paul-Sabatier, the German Academic Exchange Service (DAAD), Naturalia & Biologia (Paris) and the Institut Universitaire de France.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.G. (giurfa@cict.fr).

Structural basis of long-term potentiation in single dendritic spines

Masanori Matsuzaki¹, Naoki Honkura¹, Graham C. R. Ellis-Davies² & Haruo Kasai¹

¹Department of Cell Physiology, National Institute for Physiological Sciences and The Graduate University of Advanced Studies (Sokendai), Myodaiji, Okazaki 444-8787, Japan

²Department of Pharmacology and Physiology, Drexel University College of Medicine, Philadelphia, Pennsylvania 19102, USA

Dendritic spines of pyramidal neurons in the cerebral cortex undergo activity-dependent structural remodelling^{1–5} that has been proposed to be a cellular basis of learning and memory⁶. How structural remodelling supports synaptic plasticity^{4,5}, such as long-term potentiation⁷, and whether such plasticity is input-specific at the level of the individual spine has remained

unknown. We investigated the structural basis of long-term potentiation using two-photon photolysis of caged glutamate at single spines of hippocampal CA1 pyramidal neurons⁸. Here we show that repetitive quantum-like photorelease (uncaging) of glutamate induces a rapid and selective enlargement of stimulated spines that is transient in large mushroom spines but persistent in small spines. Spine enlargement is associated with an increase in AMPA-receptor-mediated currents at the stimulated synapse and is dependent on NMDA receptors, calmodulin and actin polymerization. Long-lasting spine enlargement also requires Ca^{2+} /calmodulin-dependent protein kinase II. Our results thus indicate that spines individually follow Hebb's postulate for learning. They further suggest that small spines are preferential sites for long-term potentiation induction, whereas large spines might represent physical traces of long-term memory.

We investigated the structural plasticity of individually identified spines on proximal apical dendrites of CA1 pyramidal neurons in rat hippocampal slice preparations, using two-photon uncaging (at 720 nm) of a caged glutamate compound (MNI (4-methoxy-7-nitroindolyl)-glutamate)⁸. Neurons within the hippocampal slices were transfected biolistically with an expression vector for enhanced green fluorescent protein (eGFP) and were imaged with a second two-photon laser at 910 nm. Two-photon uncaging allows

the release of glutamate within a small focal volume⁸, so that activation of glutamate receptors is limited to a region within $\sim 1 \mu\text{m}$ of the uncaging beam. The power of the laser was set at $\sim 5 \text{ mW}$, which induced currents through AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)-sensitive receptors with amplitudes in the range of miniature excitatory postsynaptic currents (see Methods). Repetitive stimulation of NMDA (*N*-methyl-D-aspartate)-sensitive receptors at a frequency of 1 Hz for 1 min was achieved using an external solution lacking Mg^{2+} .

This stimulation procedure induced marked increases ($>50\%$) in spine fluorescence in most spines examined (38 out of 40 spines, Fig. 1a, b; Supplementary Movie 1). The fluorescence profile (Fig. 1e) of the spine shown in Fig. 1a was indicative of a 28% increase in spine-head diameter at 81 min after stimulation, which would account for the 111% increase in fluorescence at this time (Fig. 1b). The maximal increase in spine-head volume (ΔV) was $203 \pm 37\%$ (mean $\pm$ s.e.m., $n = 40$) (Fig. 1f) and was apparent 1–5 min after stimulation. Such spine enlargement did not occur in the absence of MNI-glutamate. Furthermore, it was either transient, with spines recovering their original volume, or persistent for periods of at least 100 min. The latter fate was observed in 30% of stimulated spines and increases of more than 50% were detected for 70–100 min after stimulation. The spine enlargement was detected within 60 s of stimulation, being first apparent within 10 s of

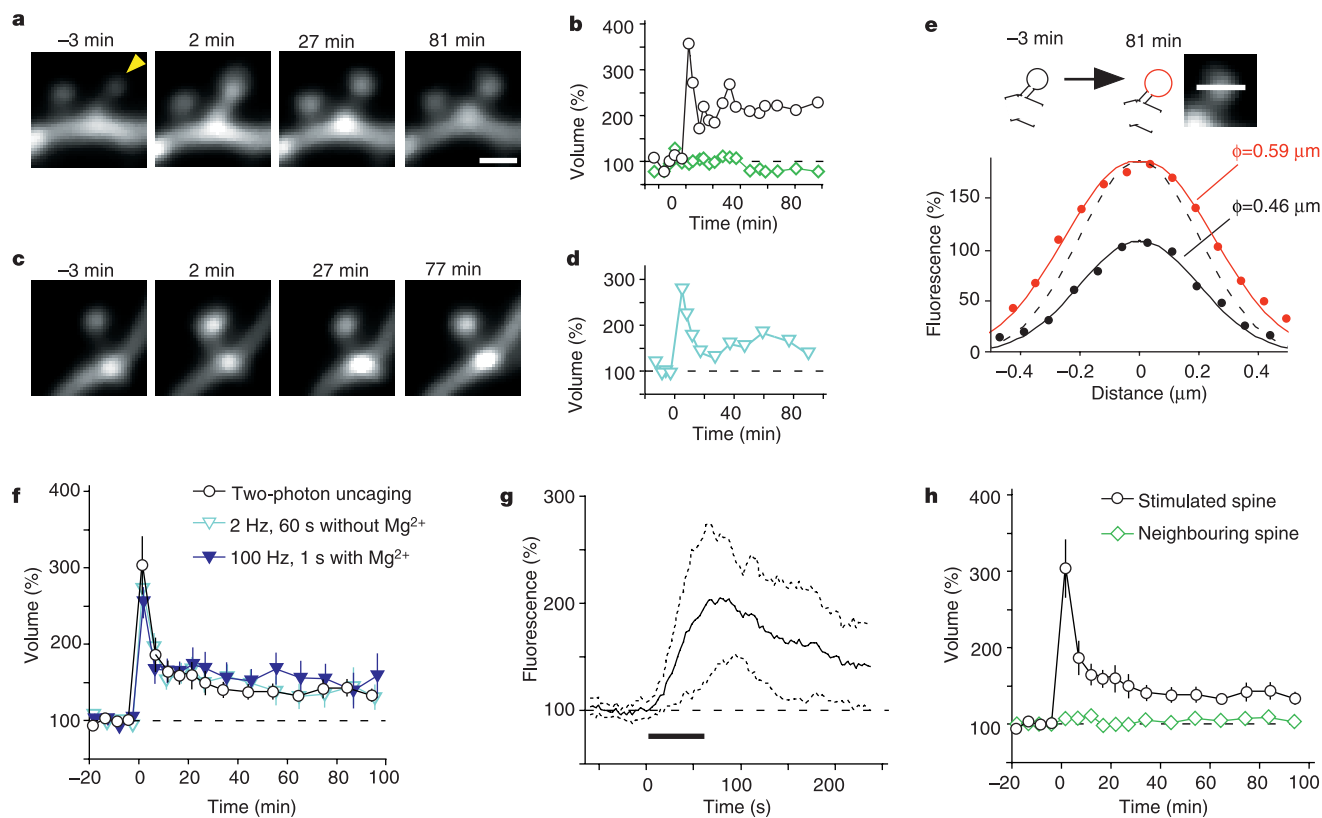


Figure 1 Spine-head enlargement induced by repetitive uncaging of MNI-glutamate. **a**, Time-lapse images of dendritic spines on a hippocampal CA1 pyramidal neuron expressing eGFP. The arrowhead indicates the spot of two-photon uncaging of MNI-glutamate, which was achieved by stimulation for 1 min at 1 Hz, beginning at time zero in a bathing solution lacking Mg^{2+} . Images were stacked along the z axis. Scale bar, $1 \mu\text{m}$. **b**, Time course of spine-head volume, estimated from stacked images, of the stimulated (black circles) and neighbouring (green diamonds) spines shown in **a**. **c**, Time-lapse images of a spine on a dendrite that was affected by electrical stimulation of presynaptic fibres at 2 Hz for 1 min; **d**, time course of the head volume. **e**, Fluorescence intensity along a line crossing the centre of the stimulated spine head shown in **a** at -3 min (black circles) and 81 min (red circles). Smooth black and red lines represent the predictions

made for the z-stacked images of spheres with diameters of 0.46 and $0.59 \mu\text{m}$, respectively (see Supplementary Methods). The dashed line is scaled from the smooth black one. **f**, **h**, Averaged time courses of spine-head volume for two-photon uncaging of MNI-glutamate (circles, $n = 40$), neighbouring spines (green diamonds in **h**, $n = 34$), 2-Hz electrical stimulation in the absence of Mg^{2+} (blue open triangles, $n = 19$) or 100-Hz electrical stimulation in the presence of Mg^{2+} (blue closed triangles, $n = 14$). Error bars represent $\pm$ s.e.m. **g**, Initial time course of spine fluorescence. The intensity in a single plane drawn through the centre of the spines was sampled; the average for 11 spines is shown. Dotted lines show time courses corresponding to the mean $\pm$ s.d., and the horizontal bar indicates the period of repetitive uncaging.

initiation of the stimulus (Fig. 1g; for such rapid imaging, the fluorescence sample had to be limited to a single plane through the centre of the spine head). The enlargement was selective towards stimulated spines (Fig. 1a, b, h); only 9% of neighbouring spines (mean distance from stimulated spines, $1.2 \mu\text{m}$; mean $\Delta V = 8 \pm 8\%$; $n = 34$) showed enlargement (of $>50\%$) and only 3% showed long-lasting enlargement.

Similar spine enlargement was also induced by repetitive stimulation of Schaffer collateral fibres, either at low frequency (2 Hz, 1 min) in the absence of Mg^{2+} (Fig. 1c, d, f; blue open triangles) or at high frequency (100 Hz for 1 s, twice with a 20-s interval) in the presence of Mg^{2+} (Fig. 1f, blue closed triangles). With such stimulation of presynaptic fibres, however, it was not possible to ascertain whether the enlarged spines corresponded to the stimulated spines. A fundamental advantage of our optical approach is that it allows the location of stimulation to be related directly to the site of spine enlargement. Given that stimulation of Schaffer collaterals and focal glutamate application induced similar structural changes in dendritic spines, we conclude that uncaging of

glutamate mimics synaptic stimulation with a resolution at the level of the individual spine.

Enlargement ($\Delta V > 50\%$) was detected in 100% of small spines (volume $< 0.1 \mu\text{m}^3$, $n = 20$) and 90% of large spines (volume $> 0.1 \mu\text{m}^3$, $n = 20$) after stimulation by uncaging of glutamate (Fig. 2a–c). The amplitude of transient enlargement ($\Delta V_{\text{transient}}$) (Fig. 2c) did not differ significantly between small and large spines (Fig. 2d). In contrast, long-lasting enlargement (ΔV_{long}) was markedly dependent on original spine size (Supplementary Fig. 2b); the mean value of ΔV_{long} was thus greater for small spines than for large ones (Fig. 2d), and long-lasting enlargement ($\Delta V_{\text{long}} > 50\%$) was observed in 55% and 5% of small and large spines, respectively (see Supplementary Information). The high-frequency electrical stimulation (Fig. 1f, 100 Hz) consistently induced long-lasting enlargement ($>50\%$) in only small spines (volume $< 0.1 \mu\text{m}^3$, $n = 5$).

The long-lasting spine enlargement showed pharmacological properties similar to those of the induction of long-term potentiation (LTP)^{9,10}. Spine enlargement was thus prevented by the NMDA receptor antagonist, D(-)-2-amino-5-phosphonovaleric acid (AP5), (Fig. 2e, f) but unaffected by the metabotropic glutamate receptor blocker, (S)-alpha-methyl-4-carboxyphenylglycine (MCPG) (Fig. 2f), indicating that it was dependent on the activation of NMDA receptors. The structural plasticity was also blocked by the calmodulin inhibitors, W7 (N-(6-amino-hexyl)-5-chloro-1-naphthalene-sulphonamide hydrochloride) (Fig. 2f) and calmidazolium (data not shown), as well as by latrunculin A (100 nM) (Fig. 2f), an inhibitor of actin polymerization, implicating Ca^{2+} - and calmodulin-dependent actin reorganization in this process. An inhibitor of Ca^{2+} /calmodulin-dependent protein kinase II, KN62, blocked the long-lasting (Fig. 2e, f) but not the transient spine enlargement. We also found that a low concentration of latrunculin A (20 nM) selectively affected the long-term but not early potentiation¹¹, and suggesting that distinct cytoskeletal organizations may underlie transient and long-lasting spine enlargements.

We next investigated expression of postsynaptic LTP during spine structural plasticity using three-dimensional mapping of AMPA receptor-mediated currents. The currents were mapped using two-photon uncaging of MNI-glutamate⁸, rather than by iontophoretic glutamate application¹². Pyramidal neurons expressing eGFP were subjected to whole-cell clamping in the perforated-patch mode to prevent washout of intracellular proteins¹³. Paired stimulation—consisting of postsynaptic depolarization to 0 mV and repetitive glutamate application to a single spine at 2 Hz for 60 s in a solution containing a physiological concentration (1 mM) of Mg^{2+} —induced structural plasticity (Fig. 3a) similar to that evoked by uncaging of MNI-glutamate in a solution lacking Mg^{2+} (Fig. 1a). Because, in the presence of Mg^{2+} , spine enlargement was observed only when depolarization was paired with glutamate application, we were able to obtain AMPA current maps⁸ (holding potential, -60 mV) both before and after the induction of structural plasticity without influencing plasticity itself. AMPA currents were markedly potentiated by paired stimulation at sites restricted to the stimulated spine (Fig. 3a, b). The potentiation was rapid, being detected as early as 3 min after stimulation (designated ‘early potentiation’; Fig. 3a). Furthermore, current potentiation ($>30\%$ increase) persisted for at least 30 min in 44% of small spines (Fig. 4b; $n = 9$). The AMPA currents were always detected at the same locations in the spines, possibly reflecting the postsynaptic density, before and after stimulation (Fig. 3). Long-lasting potentiation of AMPA currents is therefore identical to LTP in that it probably occurs at the postsynaptic density, is induced by glutamate application¹⁴ and is blocked by AP5 (see below).

Figure 4 summarizes the results from many dual imaging/mapping experiments. We found that potentiation of AMPA-receptor-

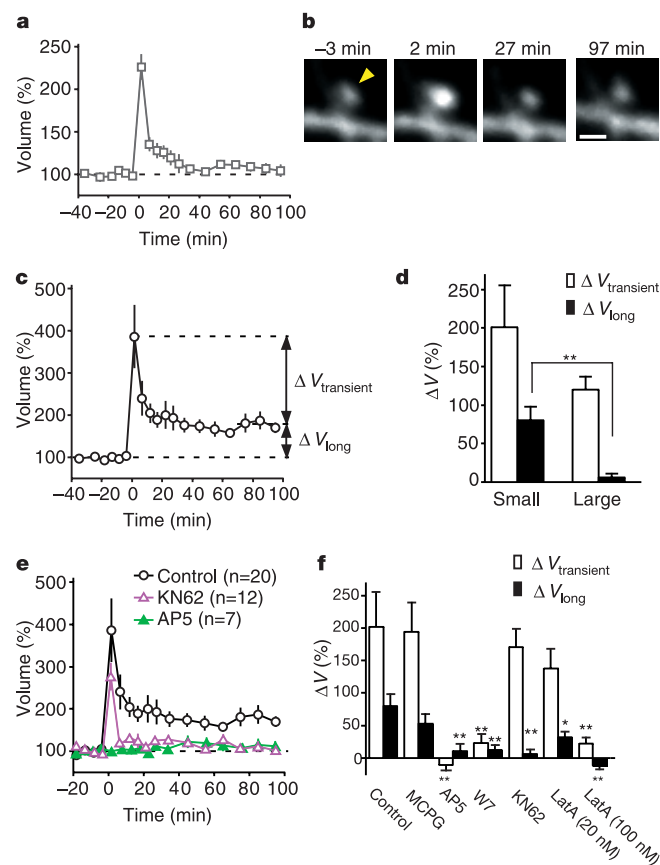


Figure 2 Properties of the spine-head enlargement. **a, c**, Averaged time course ($\pm$ s.e.m.) of spine-head volume for large (**a**) and small (**c**) spines with an initial head volume of $>0.1 \mu\text{m}^3$ ($n = 20$) and $<0.1 \mu\text{m}^3$ ($n = 20$), respectively. $\Delta V_{\text{transient}}$ represents the difference in spine-head volume between 1–5 and 70–100 min after stimulation, whereas ΔV_{long} represents that between 70–100 min after and 0–20 min before stimulation. **b**, Example of an enlargement of a large spine. The arrowhead indicates the point of two-photon uncaging of MNI-glutamate. **d**, Mean values of $\Delta V_{\text{transient}}$ and ΔV_{long} for small and large spines ($n = 20$ each). **e**, Effects of 50 μM AP5 or 4 μM KN62 on the averaged time course ($\pm$ s.e.m.) of spine-head volume for small spines. **f**, Pharmacological properties of the transient and long-lasting phases of head enlargement in small spines. Antagonists examined were 1 mM MCPG ($n = 15$), 50 μM AP5 ($n = 7$), 20 μM W7 ($n = 11$), 4 μM KN62 ($n = 12$) and 20 nM ($n = 12$) or 100 nM latrunculin A ($n = 7$). * $P < 0.05$, ** $P < 0.01$ (Mann–Whitney U test). Errors bars in **d** and **f** show s.e.m.

mediated currents at small spines was strongly correlated with enlargement of the same spine head. The following findings indicate this correlation: (1) early potentiation was detected only at spines that underwent enlargement (Fig. 4b, open circles), being apparent in seven of the 14 spines that showed enlargement (of >30%; we set the criterion level based on two s.d. values of spontaneous fluctuation; see Supplementary Information) but in neither of the two spines that did not (<30% increase in volume) (Fig. 3c, d). (2) LTP was induced in four of the five spines that showed long-lasting enlargement but in none of the four spines that did not (Fig. 4b, filled symbols). (3) The magnitude of potentiation correlated well with that of early ($\Delta V_{\text{transient}} + \Delta V_{\text{long}}$) (Fig. 4b, solid line) or long-lasting (Fig. 4b, dashed line) spine enlargement. (4) Both current potentiation and morphological enlargement were detected only in stimulated spines, not in neighbouring spines (Figs 3a, b and 4a). (5) Neither current potentiation nor spine enlargement was detected in cells in which the induction of LTP was prevented by AP5. Paired stimulation in the presence of AP5 did not substantially alter spine volume (increase of only $8 \pm 4\%$, $n = 5$) but did induce a slight decrease (of $21 \pm 9\%$, $n = 5$) in the amplitude of AMPA currents in the early phase (Fig. 4b, red cross) that was probably due to a side effect of repetitive optical illumination (similar reductions

were also induced in the following two conditions). (6) In neurons subjected to whole-cell dialysis, we detected little structural plasticity ($\Delta V = 9 \pm 8\%$, $n = 7$) and a slight decrease in AMPA currents (of $21 \pm 13\%$) (Fig. 4b, yellow cross); the absence of spine enlargement and LTP was probably due to washout of the responsible molecules¹⁵. Large spines showed little enlargement when cells were clamped in the perforated-patch mode (Figs 3d and 4b, squares in Fig. 4b; $n = 8$; see Supplementary Information), and a slight reduction in the AMPA current was detected (Fig. 4b).

We have shown that synaptic potentiation is closely related to enlargement of spine heads in terms of pharmacology, amplitude, time course and spatial localization within the dendrite. Spine enlargement was induced with little time delay, as has been reported for LTP¹⁶. In contrast, new filopodia or spines require at least 20 min to emerge from dendrites after the induction of LTP^{4,5} and therefore cannot explain the rapid onset of LTP. Under our experimental conditions, in which only one spine is stimulated, the generation of new spines was not observed. The growth of new spines might thus require the stimulation of many synapses and contribute to a later phase of LTP. The expression of AMPA receptors in spines is dependent on F-actin^{11,17} and scaffolding proteins^{18,19}. Spine enlargement, which is also dependent on actin polymerization,

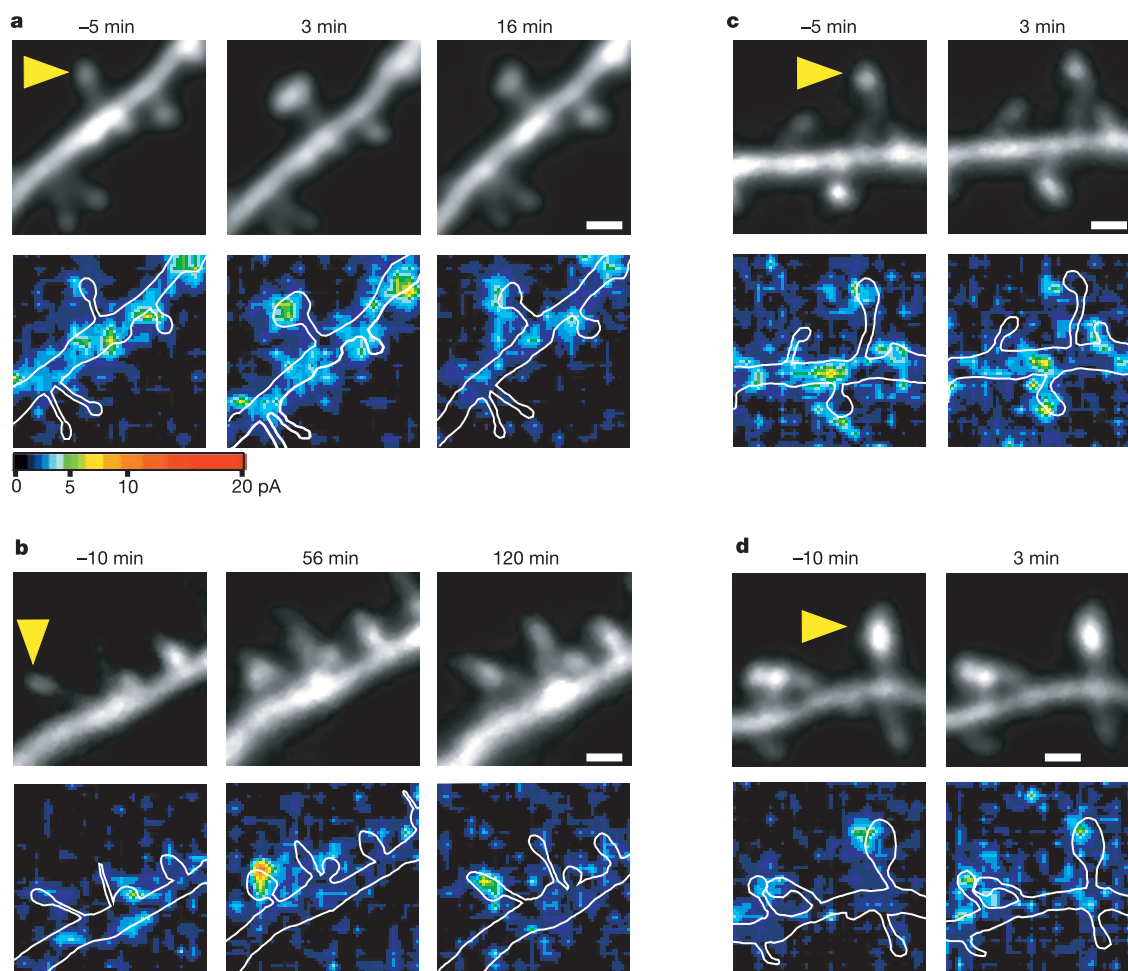


Figure 3 Colocalization of enlargement of spine heads and potentiation of AMPA-receptor-mediated currents. **a, b**, Examples of small spines that showed transient (**a**) or long-lasting (**b**) enlargement (upper panels) and potentiation of AMPA currents (lower panels). Three-dimensional mapping of AMPA currents was performed in neurons that were clamped in the perforated-patch mode. The amplitude of AMPA currents is pseudocolour coded and stacked along the z axis by the maximum-intensity method. The

neurons were depolarized to 0 mV and the small spines, indicated by the arrowheads, were stimulated by two-photon uncaging of MNI-glutamate at 2 Hz between times 0 and 60 s. White lines in the lower panels indicate contours of dendrites. Scale bars, 1 μm. **c, d**, Examples of small (**c**) and large (**d**) spines whose heads failed to enlarge in response to paired stimulation.

might therefore promote the accumulation of AMPA receptors. However, the precise molecular mechanisms of F-actin reorganization^{20,21} and AMPA receptor expression^{18,19,22,23} remain to be clarified.

With the use of location-specific two-photon photolysis of caged glutamate to stimulate single synapses, we have shown that both

structural and functional plasticity can be induced at the level of the individual spine. Our results thus indicate that a hebbian mechanism²⁴ can operate individually in single spines of hippocampal CA1 pyramidal neurons as has been suggested²⁵ but not proved²⁶. We also found that the preferential sites of LTP induction are small spines, which express a small number of AMPA receptors⁸. Small spines might thus correspond to the postsynaptic structures of so-called silent synapses^{27–29}. Given that large spines have been shown to express AMPA receptors abundantly⁸ and to be stable for months in the mouse cerebral cortex *in vivo*^{2,3}, and given that we have now shown that small spines can be converted into larger spines, we suggest that spines act as memory units, with large spines being the physical traces of long-term memory. Consistent with this idea, we found that large spines are resistant to LTP, which is critical for information storage. The memory stored in large spines might thus be protected from further potentiation caused by its readout and new memory formation³⁰. Further clarification of the structural bases of synaptic potentiation and depression should give greater insight into the operation of memory units in the brain. □

Methods

Preparation of slice cultures

Hippocampal slices with a thickness of 350 μm were prepared from 6–8-day-old Sprague–Dawley rats. The slices were mounted on 0.4- μm culture inserts (Millipore) and incubated at 35 °C under 5% CO_2 in a medium comprising 50% MEM (Invitrogen), 25% Hanks' balanced salt solution (Invitrogen), 25% horse serum (Nichirei) and 6.5 g l^{-1} glucose. After 5–6 days *in vitro*, slices were transfected using a gene gun system (PDS-1,000; Bio-Rad), with a vector containing eGFP complementary DNA under the control of the chicken β -actin gene promoter and cytomegalovirus enhancer. Two to five days after transfection, each slice was transferred to a recording chamber and superfused at room temperature (23–25 °C) with a solution that contained 125 mM NaCl, 2.5 mM KCl, 3 mM CaCl_2 , 1.25 mM NaH_2PO_4 , 26 mM NaHCO_3 and 20 mM glucose, and which was bubbled with 95% O_2 and 5% CO_2 . The bathing solution also contained 1 μM tetrodotoxin, 50 μM picrotoxin and 200 μM Tetroxol (Aldrich). For electrical stimulation, 0.1-ms voltage pulses were delivered either with a glass pipette in the absence of Mg^{2+} , or with a bipolar tungsten electrode in the presence of Mg^{2+} . In experiments with the antagonists AP5, MCPG (Tocris), W7 (Calbiochem), KN62 (Seikagaku) and latrunculin A (Molecular Probes), slices were superfused with bathing solution containing each drug at least 30 min before two-photon uncaging of caged glutamate. The experiments were approved by the animal experiment committee of the National Institute for Physiological Sciences.

Two-photon excitation imaging and uncaging

Time-lapse two-photon imaging of dendritic spines was performed with an upright microscope (BX50WI; Olympus) equipped with a water immersion objective lens (LUMPlanFI/IR $\times 60$, numerical aperture of 0.9; Olympus) and with Fluoview software (Olympus). Two mode-locked fs-pulse Ti:sapphire lasers (Spectra Physics) set at wavelengths of 720 and 910 nm were connected to the laser-scanning microscope through two independent scanheads. For imaging of dendritic spines we used the laser set at a wavelength of 910 nm, with which the point-spread function of the focal volume was estimated using 0.1- μm fluorescent beads as 0.44 μm (full-width at half-maximum) laterally and 2.1 μm axially. For three-dimensional reconstruction, 13–17 xy images separated by 0.5 μm were stacked by summation of fluorescence values at each pixel, thus the enlargements appear as an increase in diameter and intensity. Immediately before two-photon uncaging, one image was acquired at the 720-nm wavelength for precise adjustment of the point of photolysis; MNI-glutamate (12 mM) was then applied locally from a glass pipette positioned close to the selected dendrite. Repetitive (1 or 2 Hz) photolysis of MNI-glutamate was performed on the specimen at a 720-nm wavelength with a pulse-train duration of 0.6 ms and a power of ~ 5 mW. A self-made program based on LabView (National Instruments) controlled a galvano-scanner driver and a mechanical shutter (Uniblitz).

Electrophysiology

For perforated-patch recordings, 1 mM MgCl_2 was added to the extracellular solution and the concentration of CaCl_2 was increased to 4 mM. The patch-clamp electrodes (open-tip resistance, 1.5–2.5 M Ω) were filled with a solution containing 136.5 mM potassium gluconate, 17.5 mM KCl, 9 mM NaCl, 1 mM MgCl_2 , 0.2 mM EGTA (tetrapotassium salt), 10 mM HEPES-KOH (pH 7.2) and 0.3 mg ml^{-1} amphotericin B. Neurons were voltage-clamped at -60 mV and the currents were low-pass filtered at 2 kHz and sampled at 10 kHz. Series resistance was 32 ± 8.1 M Ω (mean $\pm$ s.d.) during mapping of AMPA receptors. Amplitudes of AMPA currents were 1–25 pA, (measured with perforated patch-clamp experiments), which were within the range of miniature excitatory postsynaptic currents⁸. For such mapping, a pseudorandom sequence of scanning pixels in a region of interest was constructed to maintain the distance between two successive pixels at greater than 2.5–5 μm and AMPA currents were sampled at an interval of 100 ms. For three-dimensional mapping of AMPA currents, xy scanning of a region of interest comprising 16 $\times$ 16 pixels (1 pixel = 0.33 μm) was performed at 4–6 different z axis planes, each separated by 1 μm . Peak amplitudes of AMPA currents were assigned to each pixel, linear-

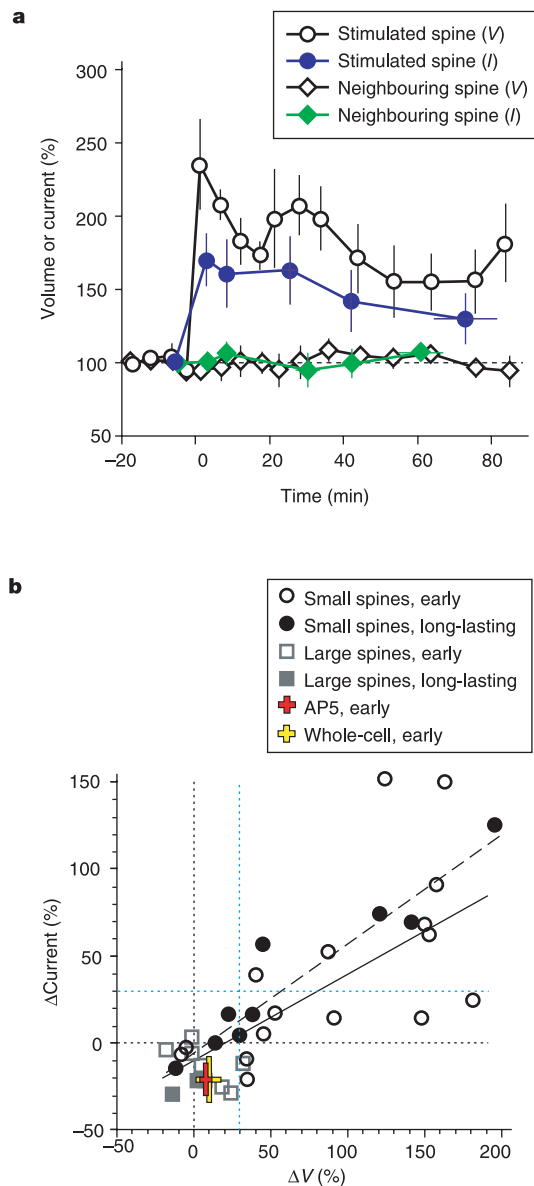


Figure 4 Relationship between spine-head enlargement and potentiation of AMPA-receptor-mediated currents. **a**, Time courses ($\pm$ s.e.m.) of spine-head volume (V, open symbols) and maximal AMPA currents (I, filled symbols) normalized to the original values. The data were derived from all the small spines ($n = 9$) that showed enlargement immediately after pairing stimulation and for which it was possible to obtain current maps more than 30 min after paired stimulation in the perforated-patch recording mode. Circles and diamonds represent stimulated and neighbouring spines, respectively. **b**, Relationships between normalized amplitudes of head enlargement (ΔV) and potentiation of AMPA currents (Δ Current) in the early phase (1–5 min after stimulation) and long-lasting phase (30–120 min) in small and large spines. The continuous and dashed lines are regression lines for the early and long-lasting phases, respectively; correlation coefficients are 0.72 ($n = 24$, $P < 0.0001$) and 0.95 ($n = 9$, $P < 0.0001$), respectively. The red and yellow crosses show data obtained in the presence of AP5 ($n = 5$) or with whole-cell recording ($n = 7$), respectively. Black and blue dotted lines are 0% and 30% levels, respectively.

interpolated⁸, stacked along the z axis by the maximum-intensity method and displayed with a pseudocolour coding (Fig. 3a).

Received 18 December 2003; accepted 10 May 2004; doi:10.1038/nature02617.

Published online 9 June 2004.

1. Van Harrevel, A. & Fikova, E. Swelling of dendritic spines in the fascia dentata after stimulation of the perforant fibers as a mechanism of post-tetanic potentiation. *Exp. Neurol.* **49**, 736–749 (1975).
2. Trachtenberg, J. T. *et al.* Long-term *in vivo* imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* **420**, 788–794 (2002).
3. Grutzendler, J., Kasthuri, N. & Gan, W. B. Long-term dendritic spine stability in the adult cortex. *Nature* **420**, 812–816 (2002).
4. Maletic-Savatic, M., Malinow, R. & Svoboda, K. Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. *Science* **283**, 1923–1927 (1999).
5. Engert, F. & Bonhoeffer, T. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* **399**, 66–70 (1999).
6. Yuste, R. & Bonhoeffer, T. Morphological changes in dendritic spines associated with long-term synaptic plasticity. *Annu. Rev. Neurosci.* **24**, 1071–1089 (2001).
7. Bliss, T. V. & Lomo, T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol. (Lond.)* **232**, 331–356 (1973).
8. Matsuzaki, M. *et al.* Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons. *Nature Neurosci.* **4**, 1086–1092 (2001).
9. Malenka, R. C. *et al.* An essential role for postsynaptic calmodulin and protein kinase activity in long-term potentiation. *Nature* **340**, 554–557 (1989).
10. Malinow, R., Schulman, H. & Tsien, R. W. Inhibition of postsynaptic PKC or CaMKII blocks induction but not expression of LTP. *Science* **245**, 862–866 (1989).
11. Krucker, T., Siggins, G. R. & Halpain, S. Dynamic actin filaments are required for stable long-term potentiation (LTP) in area CA1 of the hippocampus. *Proc. Natl Acad. Sci. USA* **97**, 6856–6861 (2000).
12. Davies, S. N., Lester, R. A., Reymann, K. G. & Collingridge, G. L. Temporally distinct pre- and post-synaptic mechanisms maintain long-term potentiation. *Nature* **338**, 500–503 (1989).
13. Horn, R. & Marty, A. Muscarinic activation of ionic currents measured by a new whole-cell recording method. *J. Gen. Physiol.* **92**, 145–159 (1988).
14. Cormier, R. J., Mauk, M. D. & Kelly, P. T. Glutamate iontophoresis induces long-term potentiation in the absence of evoked presynaptic activity. *Neuron* **10**, 907–919 (1993).
15. Kato, K., Clifford, D. B. & Zorumski, C. F. Long-term potentiation during whole-cell recording in rat hippocampal slices. *Neuroscience* **53**, 39–47 (1993).
16. Gustafsson, B. & Wigstrom, H. Long-term potentiation in the hippocampal CA1 region: its induction and early temporal development. *Prog. Brain Res.* **83**, 223–232 (1990).
17. Kim, C. H. & Lisman, J. E. A role of actin filament in synaptic transmission and long-term potentiation. *J. Neurosci.* **19**, 4314–4324 (1999).
18. Schnell, E. *et al.* Direct interactions between PSD-95 and stargazin control synaptic AMPA receptor number. *Proc. Natl Acad. Sci. USA* **99**, 13902–13907 (2002).
19. Rumbaugh, G., Sia, G. M., Garner, C. C. & Huganir, R. L. Synapse-associated protein-97 isoform-specific regulation of surface AMPA receptors and synaptic function in cultured neurons. *J. Neurosci.* **23**, 4567–4576 (2003).
20. Fischer, M., Kaech, S., Knutti, D. & Matus, A. Rapid actin-based plasticity in dendritic spines. *Neuron* **20**, 847–854 (1998).
21. Pak, D. T., Yang, S., Rudolph-Correia, S., Kim, E. & Sheng, M. Regulation of dendritic spine morphology by SPAR, a PSD-95-associated RapGAP. *Neuron* **31**, 289–303 (2001).
22. Hayashi, Y. *et al.* Driving AMPA receptors into synapses by LTP and CaMKII: requirement for GluR1 and PDZ domain interaction. *Science* **287**, 2262–2267 (2000).
23. Lisman, J., Schulman, H. & Cline, H. The molecular basis of CaMKII function in synaptic and behavioural memory. *Nature Rev. Neurosci.* **3**, 175–190 (2002).
24. Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
25. Yuste, R. & Denk, W. Dendritic spines as basic functional units of neuronal integration. *Nature* **375**, 682–684 (1995).
26. Engert, F. & Bonhoeffer, T. Synapse specificity of long-term potentiation breaks down at short distances. *Nature* **388**, 279–284 (1997).
27. Durand, G. M., Kovalchuk, Y. & Konnerth, A. Long-term potentiation and functional synapse induction in developing hippocampus. *Nature* **381**, 71–75 (1996).
28. Liao, D., Hessler, N. A. & Malinow, R. Activation of postsynaptically silent synapses during pairing-induced LTP in CA1 region of hippocampal slice. *Nature* **375**, 400–404 (1995).
29. Isaac, J. T., Nicoll, R. A. & Malenka, R. C. Evidence for silent synapses: implications for the expression of LTP. *Neuron* **15**, 427–434 (1995).
30. Kasai, H., Matsuzaki, M., Noguchi, J., Yasumatsu, N. & Nakahara, H. Structure–stability–function relationships of dendritic spines. *Trends Neurosci.* **26**, 360–368 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Haydon for reading the manuscript, N. Takahashi and T. Kise for technical assistance and Y. Yanagawa and M. Okabe for the eGFP construct. This work was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (H.K. and M.M.) and by the Human Frontier Science Program Organization (G.C.R.E.-D. and H.K.), the NIH (G.C.R.E.-D. and H.K.), the NSF (G.C.R.E.-D.) and the McKnight Endowment Fund for Neuroscience (G.C.R.E.-D.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.K. (hkasai@nips.ac.jp).

Mechano-electrical transduction of adult outer hair cells studied in a gerbil hemicochlea

David Z. Z. He^{1,2}, Shuping Jia¹ & Peter Dallos³

¹Hair Cell Biophysics Laboratory, Department of Biomedical Sciences, Creighton University, Omaha, Nebraska 68175, USA

²Boys Town National Research Hospital, Omaha, Nebraska 68131, USA

³Auditory Physiology Laboratory (The Hugh Knowles Center), Depts. of Neurobiology and Physiology, and Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60208, USA

Sensory receptor cells of the mammalian cochlea are morphologically and functionally dichotomized. Inner hair cells transmit auditory information to the brain, whereas outer hair cells (OHC) amplify the mechanical signal, which is then transduced by inner hair cells¹. Amplification by OHCs is probably mediated by their somatic motility^{2,3} in a mechanical feedback process. OHC motility *in vivo* is thought to be driven by the cell's receptor potential. The first steps towards the generation of the receptor potential are the deflection of the stereociliary bundle⁴, and the subsequent flow of transducer current through the mechanosensitive transducer channels located at their tips⁵. Quantitative relations between transducer currents and basilar membrane displacements are lacking, as well as their variation along the cochlear length. To address this, we simultaneously recorded OHC transducer currents (or receptor potentials) and basilar membrane motion in an excised and bisected cochlea, the hemicochlea⁶. This preparation permits recordings from adult OHCs at various cochlear locations while the basilar membrane is mechanically stimulated. Furthermore, the stereocilia are deflected by the same means of stimulation as *in vivo*. Here we show that asymmetrical transducer currents and receptor potentials are significantly larger than previously thought, they possess a highly restricted dynamic range and strongly depend on cochlear location.

Until now, transducer currents have been recorded from solitary mammalian OHCs (ref. 7), organotypic cultures from immature ears⁸ or excised segments of sensory epithelia⁹ by directly deflecting the cells' stereocilia. We recorded OHC transducer currents (or receptor potentials) from a more mechanical *in vivo*-like preparation, the gerbil hemicochlea (Fig. 1a). This preparation allowed us to correlate transducer current (or receptor potential) and basilar membrane motion in adult animals. The hemicochlea is a highly specialized preparation; it is a cochlea cut in half, hence it does not have normal hydrodynamic properties. Furthermore, as with any *in vitro* preparation, it does not sustain normal cochlear amplification. The reason is thought to be that the presumed means of amplification, OHC electromotility, is smaller *in vitro* because of a reduction in electric driving force. This reduction is because, *in vitro*, the endocochlear potential is absent and the OHC resting potential is reduced. Nevertheless, the hemicochlea is an ideal preparation for examining mechano-electrical properties of hair cells *in situ*, with the hair cells being stimulated through their usual micro-mechanical environment⁵. Even if the surrounding OHCs do sustain some degree of electromotility (as they probably do, maintained by voltage changes around their reduced resting potential) the cell from which transducer currents are measured will not do so, as it is voltage-clamped and electromotility is a voltage-driven process. Conversely, when we measure receptor potentials, the cells are not voltage clamped and probably have some motility.

Gerbils ranging in age between 25 and 30 days old were used. Transducer currents were recorded using the tight-seal whole-cell

voltage-clamp technique (Fig. 1b). The basilar membrane displacement shows shallow and broad tonotopic tuning in the hemicochlea⁶. We measured transducer currents and basilar membrane displacement from 46 apical- and 21 basal-turn OHCs. The approximate recording locations were 2.7–3.2 mm and 8.6–8.9 mm from the base of the cochlea, corresponding to 18–13 kHz and 1–0.6 kHz best frequencies *in vivo*¹⁰. As 102 Hz sinusoidal commands were used to evoke basilar membrane vibration, this stimulating frequency was well below the characteristic frequencies of the measuring locations.

Examples of simultaneous recordings of the current and basilar membrane displacement from third-row OHCs in the apical and basal turns are shown (Fig. 2a, b). The basilar membrane displacements in both turns were essentially symmetrical, and linear in magnitude within the range measured. The transducer currents at a holding potential of -70 mV were, however, highly asymmetric; basilar membrane motion towards the scala vestibuli produced a

large increase in inward current, whereas basilar membrane displacement towards the scala tympani yielded a very small decrease in inward current (segments of the current and basilar membrane responses are plotted on an expanded time scale, see insert Fig. 2a). This confirms the general assumption that upward motion of the basilar membrane towards the scala vestibuli pushes the hair bundle in the direction of the tallest stereocilia, and thus opens transducer channels. Strong asymmetry was observed in about 75% of the cells measured. Less asymmetrical responses were observed in the remainder of cells (data not shown). Highly asymmetrical responses occurred more frequently in later experiments, presumably because of our greater facility with the preparation. Consequently, we assumed that more symmetrical responses signified sub-optimal conditions of the hemicochlea. The results were compared to extensive data on the direct stimulation of the stereociliary bundle^{4,8}. Displacements that closed transducer channels produced current that showed saturation even when the basilar membrane displacement (towards the scala tympani) was very small. In contrast, in most cases no clear saturation in the opposite direction (towards the scala vestibuli) was observed, even at the maximal basilar membrane displacement used. The use of larger mechanical stimuli was avoided in order to prevent the loss of the seal during patch recordings. In Fig. 2c and d we show two examples of responses that reached saturation. One of these is from a basal-turn cell (Fig. 2c) for which the basilar membrane displacement was larger. Assuming that the paddle vibration was the same in all cases, larger motions corresponded to more compliant cochlear partitions. Because in this case the saturating current response was also low, we assumed that damage was inflicted on the preparation. A fully-saturated response from an apical-turn cell is also shown (Fig. 2d). In Fig. 2e and f magnitudes of transducer currents, estimated conductances and slopes of the current plots are presented as a function of basilar membrane displacement for two of our best, albeit not fully saturated, cells. Current data points were fitted with second order Boltzmann functions, and slope plots with their derivative (solid lines). The maximum measurable conductances at -70 mV for the two examples were 32.4 nS (basal) and 12.4 nS (apical). In order to estimate fully-saturating inward current and conductance, we extrapolated the Boltzmann-fit of the current-displacement functions. The resulting values were 2.43 nA (34.7 nS) for the basal cell, and 1.18 nA (17.3 nS) for the apical cell. These values from adult OHCs are considerably in excess of those obtained from immature mouse and rat apical turn OHCs *in vitro*: 9.4 nS (refs 8, 9). Maximum slopes were 50 pA nm^{-1} and 8.2 pA nm^{-1} for the two cells. Differences in current magnitude between the cells from the basal and apical turns partly result from the difference in the number of stereocilia and, consequently, the number of transducer channels. In all known species, basal turn OHCs have more stereocilia than their apical turn counterparts. In the turtle, a single mechano-transducer channel conductance ranges between 80 and 160 pS, depending on cochlear location¹¹. It is possible that mammalian hair cells also exhibit such gradation. Thus, changing unitary conductance and the number of transducer channels along a cochlear gradient are together likely to account for the twofold change in transducer conductance demonstrated above.

The cells shown in Fig. 2e and f had approximately 0.64 nS (base) and 0.57 nS (apex) transducer conductance at rest, corresponding to an estimated 1.8% and 3.3% of the transducer channels open in base and apex, respectively. *In vivo* receptor potential recordings with sharp electrodes from OHCs are probably so 'leaky' that they represent the properties of receptor currents. Such recordings at very low frequencies demonstrated receptor potentials asymmetrical in the hyperpolarizing direction both in the base^{12,13} and the apex¹⁴. Such asymmetry was never observed in our recordings. This difference between hemicochlea and *in vivo* results is considered further below.

Finally, we call attention to the apparent sensitivity difference

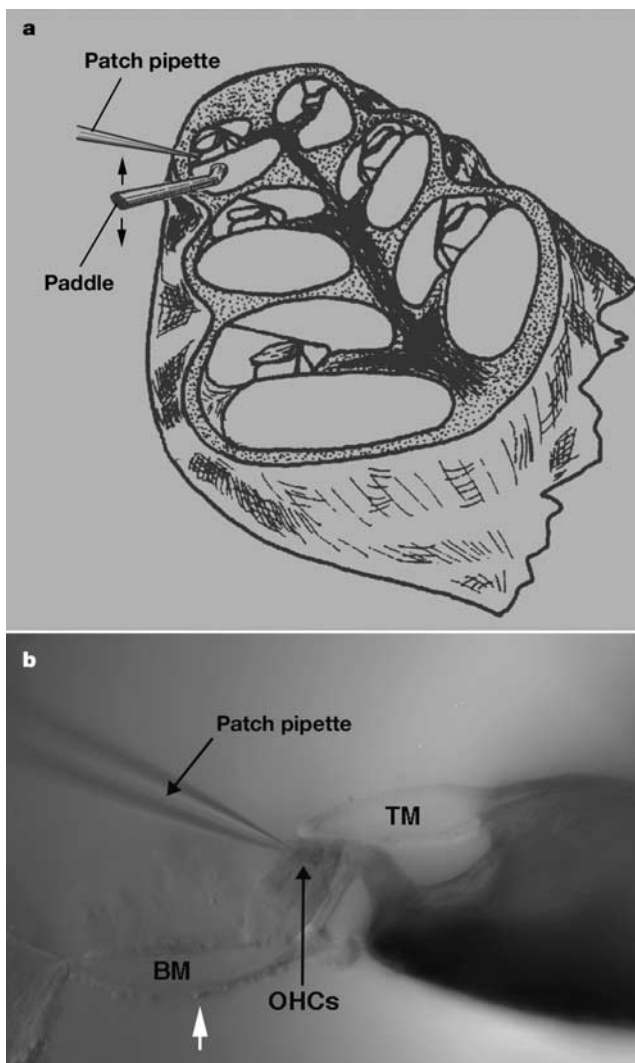


Figure 1 Recording from the hemicochlea. **a**, Schematic drawing of a hemicochlea showing the cochlear partition and the driving paddle, placed close to the mid-point of the pectinate zone of the basilar membrane in scala tympani. Double arrows represent directions of the paddle movement. **b**, Video image of whole-cell tight-seal recording of transducer current (or receptor potential) and basilar membrane motion from a third-row OHC in the low apical turn of a hemicochlea prepared from a 30-day-old gerbil. A white arrow indicates the location where the basilar membrane (BM) displacement was measured. TM: tectorial membrane.

between basal and apical cells. As seen in Fig. 2e and f, maximum sensitivity of the apical OHC was about 8.2 pA nm^{-1} , contrasted with about 50 pA nm^{-1} for the basal cell; giving a ratio of six. It is possible that two factors conspire to produce this differential. First, as considered above, a factor of two is likely because of the ratio of transducer conductances. Second, there is an estimated fourfold difference in the mechanical advantage between basilar membrane displacement to cilia angular rotation¹⁵. Our computations based on the geometry of the organ of Corti and kinematics suggested that for small signals, the ratios of cilia angle and basilar membrane displacement are 40 and $10 \text{ deg } \mu\text{m}^{-1}$ for basal and apical OHCs, respectively. Thus the expected sensitivity ratio was about eight, commensurate with the observed number of six. The estimated angular sensitivity for basal and apical hair cells was $1,250 \text{ pA deg}^{-1}$ (17.9 nS deg^{-1}) and 205 pA deg^{-1} (2.9 nS deg^{-1}). On the basis of these computations, we could further estimate basilar membrane displacement and cilia-angle ranges between 10% and 90% activation of transducer conductance. The resulting numbers were 52 nm (2.1 deg) for the base and 190 nm (1.9 deg) for the apex. Thus, the physiological range of ciliary rotation seems constant

throughout the cochlea and similar to that reported previously¹⁶. It is possible that differences between basal and apical OHCs shown here, underlie the changing gain of the cochlear amplifier for different stimulus frequencies. Judging from sensitivity data in prestin knockout mice^{3,17}, the gain changes from approximately 45 dB at low frequencies to approximately 60 dB at high frequencies.

In OHCs, the displacement of motility-related charge is controlled by both voltage and membrane tension, consequently either can evoke gating currents¹⁸. A voltage-dependent gating current was not expected as the cells were voltage clamped. Tension-dependent currents reach a peak magnitude between holding potentials of -60 and -10 mV (ref. 18). In order to ascertain that the currents recorded were not due to the gating currents of the motor protein prestin¹⁹, we measured the amplitude and reversal potential of the transducer current at various holding potentials, with the same mechanical stimulus. The amplitude of the transducer current decreased notably from the holding potential of -140 to 0 mV (Fig. 3a). It reached a minimum between 0 and $+4 \text{ mV}$, before increasing again at more positive potentials. The average reversal potential measured from five cells was $+3.6 \text{ mV}$ ($\pm 1.8 \text{ mV s.d.}$),

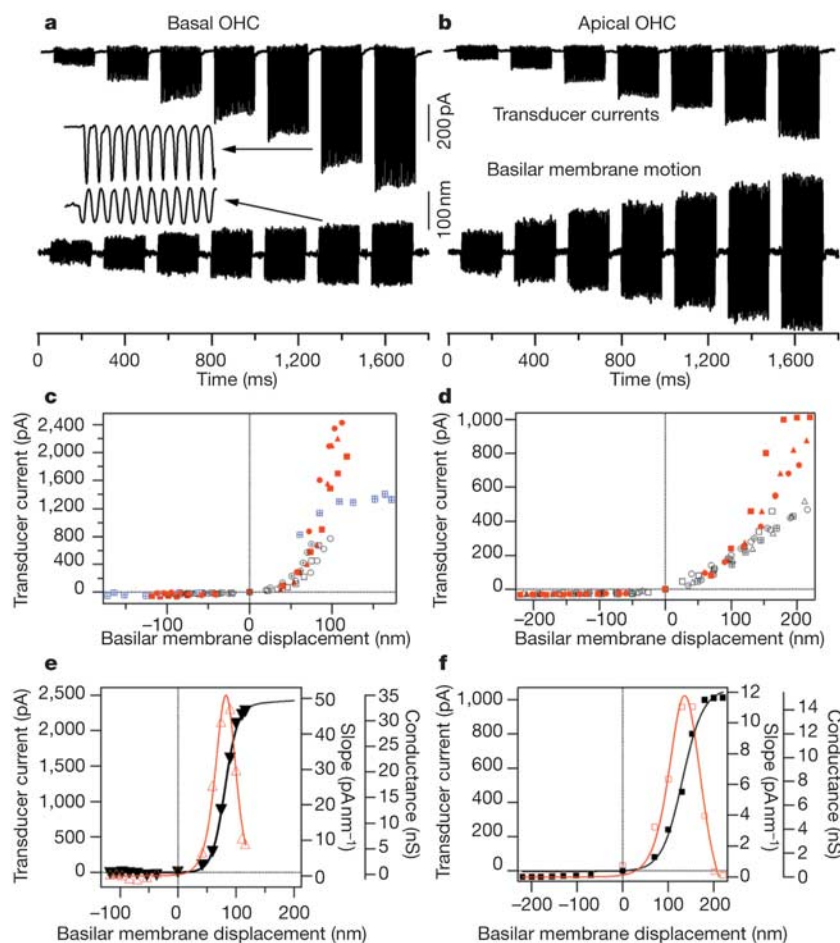


Figure 2 OHC transducer currents evoked by basilar membrane vibration. **a, b**, Tight-seal recordings were made from two OHCs in the basal and apical turns at a holding potential of -70 mV . As shown on an expanded time-scale in the insert, upward movement represents the basilar membrane moving towards the scala media, which evoked an increase in inward current (plotted downward in panels **a** and **b**). **c, d**, Steady-state transducer current plotted as a function of basilar membrane displacement for eight basal (**c**) and eight apical (**d**) OHCs. Increase in inward current is plotted upward. Responses showing the most commonly observed behaviour are plotted with black open symbols.

Some of our best cells (filled red symbols) showed larger responses and intimate saturation. One fully saturated basal-turn cell, having small response and broad dynamic range, is also shown for comparison (blue symbols). **e, f**, Transducer current, transducer conductance and slope of the current function plotted for one good basal (**e**) and apical cell (**f**). The current plots are fitted with second-order Boltzmann functions, which are shown including the extrapolated saturation region (filled symbols). Slope functions (open symbols) are fitted with the derivatives of the second-order Boltzmann functions.

similar to that found in mouse OHCs (ref. 8). Current reversal near 0 mV supports earlier data that the transducer channel is non-specific^{8,20}. Transducer currents were also measured before and after prestin's function was eliminated by removing intracellular chloride²¹. No notable reduction in current amplitude was found after this manipulation ($n = 5$, data not shown). Collectively, these results suggest that the current recorded during basilar membrane vibration in a voltage-clamped cell receives insignificant contribution, if any, from the piezo-electrical properties of the motor proteins. We also examined the transducer current after streptomycin (100 μ M), a known mechano-transducer channel blocker, was applied locally to the stereocilia. Approximately 90% of the current was blocked after streptomycin was applied, and the effect was reversible after the drug was removed (Fig. 3c). We also studied transducer conductance when the calcium content of the bathing medium was reduced to endolymph-like levels, approximately 20 μ M. In conformation with previous findings^{9,22}, we saw an increase of transducer current and conductance by at least threefold (Fig. 3d). This behaviour is presumably due to the removal of the calcium block of the transducer channels.

Receptor potentials have been recorded from guinea pig OHCs *in vivo* with acoustic stimulation^{12–14,23} and from cultured mouse OHCs in response to direct displacement of the stereocilia^{13,24}. We measured receptor potentials and receptor currents in response to direct basilar membrane displacement *in situ* using whole-cell recording techniques without channel blockers. We wondered how the receptor potential might be shaped by voltage-dependent conductances^{25,26}. In Fig. 4, an example of such measurements from

an apical turn OHC is shown. The basilar membrane displacement was essentially symmetrical, and linear in amplitude, as seen in Fig. 2a and b. The receptor current at -70 mV was highly asymmetrical and showed slow adaptation, especially at high levels, with an approximate time constant of 41 ms (ref. 27). With the sinusoidal stimuli used here, fast adaptation⁹ could not be observed. The I–V curve measured from the same cell is plotted in the insert; it is similar to those previously reported for OHCs (ref. 28). The cell's slope conductance at -70 mV was 7 nS. Both current and voltage responses have alternating current (a.c.) (cycle-by-cycle) and direct current (d.c.) (unidirectional) components. In this example, and in all cells studied, the d.c. component was always in the depolarization direction at all levels. Moreover, the d.c. shift was large enough in many cases that responses on the scala-tympani-directed half-cycle of the stimulus did not reach below the baseline (Fig. 4a, b); a behaviour not observed in transducer currents (Fig. 2). These results are different from those obtained from *in vivo* recordings, that showed, when the stimulus frequency was below the characteristic frequency, the d.c. was in the hyperpolarization direction at low levels and then switched to the depolarization direction at high levels^{12,14,29}. It is possible that the hyperpolarizing d.c. response observed *in vivo* was related to the characteristics of the travelling wave, specifically, it may be that different d.c. biases on the ciliary bundle (transducer operating point) are produced at or below best

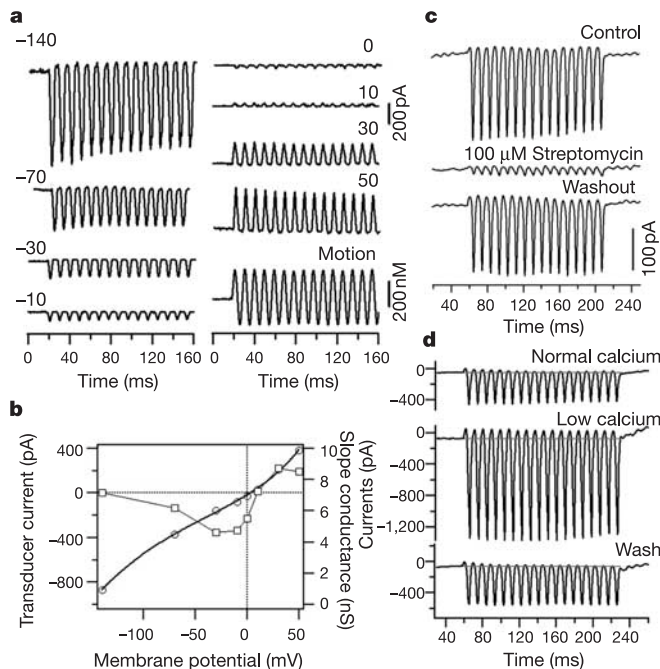


Figure 3 Recording of transducer currents. **a**, Size and polarity of transducer currents at different holding potentials (indicated by each trace) from an apical-turn OHC. The polarity reverses around 4 mV. **b**, Magnitude of the steady-state current response (circles) measured from **a**. The fit through the data are according to a simple single-energy-barrier model⁹. The slope conductance (squares) is also shown. **c**, Transducer currents of an apical OHC recorded without and with 100 μ M streptomycin applied to the cilia, and after washout of the drug. **d**, Transducer currents recorded with changing calcium concentration in the medium. The middle panel shows results with reduced calcium concentration (estimated at 10–20 μ M).

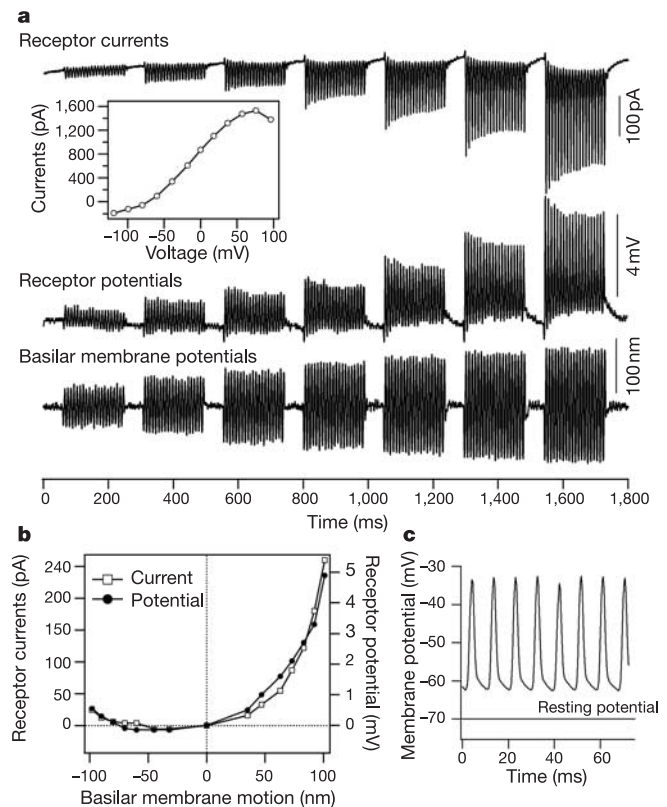


Figure 4 Recording of receptor potentials. **a**, Receptor current and receptor potential recorded from an apical turn OHC at the membrane potential of -70 mV during basilar membrane vibration. Normal pipette medium with 140 mM KCl was used. The input resistance measured at -70 mV was 156 M Ω . I–V curve obtained from whole-cell current evoked by voltage steps is shown in the insert. **b**, Receptor current and receptor potential as a function of basilar membrane displacement (measured from **a**). **c**, Another example (from an apical OHC) of receptor potential plotted on an expanded time scale. Note that the peak depolarizations are to -32 mV, representing a voltage-swing of 38 mV (peak-to-peak a.c. response is approaching 30 mV).

frequency and at different levels. Indeed, we have shown in these experiments that biasing the basilar membrane towards the scala vestibuli can reverse the polarity of the d.c. receptor current (data not shown).

For small signals, the sensitivity of the a.c. response is approximately $30 \mu\text{V nm}^{-1}$ at the holding potential of -70 mV . This value may be compared with that estimated from *in vivo* recordings from the apex, where the membrane potential is also -70 mV^{29} , or approximately 100 mV Pa^{-1} pressure at the eardrum^{14,29}, corresponding to approximately $1,000 \text{ nm Pa}^{-1}$ basilar membrane motion³⁰ in the apex, altogether yielding approximately $100 \mu\text{V nm}^{-1}$ sensitivity. Clearly, and not surprisingly, the *in vivo* preparation, possessing an active cochlear amplifier, is more sensitive than the passive hemicochlea. In our preparation, a.c. peak-to-peak responses greater than 20 mV were occasionally observed. The largest response encountered at the saturation level was approximately 30 mV . This is in agreement with the largest a.c. response recorded from guinea-pig apical-turn OHCs *in vivo*²⁹. The receptor potential in our preparation is not directly comparable to its *in vivo* counterpart as the driving force is only 70 mV , as opposed to 140 – 150 mV *in vivo*, and there is no endolymph–perilymph differential (including low calcium outside the transducer channels). As shown in Fig. 3a, changing the membrane potential to -140 mV increased the size of the transducer current by a factor of at least two. If the temperature difference and endolymph versus perilymph environment for the transducer channels are also taken into consideration, the receptor potential could be twice the size of the maximum 30 mV that was observed in our preparation. The result intimates that the a.c. responses measured *in vivo*, using sharp microelectrodes, may have underestimated the receptor potential and, in fact, at least below the cells' basolateral membrane electrical filter cut-off frequencies, these potentials are large enough to activate voltage-dependent conductances of OHCs to shape the cell's response. □

Methods

Gerbils ranging in age between 25 and 30 days old were used. Details of preparing hemicochleae have previously been described⁸. In brief, the medial surface of the cochlea was glued onto the surface of a small metal block in the cutting bath of a standard vibratome (TPI). The cochlea was aligned with the cutting edge of the blade until all cochlear-turn boundaries were perpendicular to the blade in the L-15 medium (Invitrogen). The cochlea was sliced into two parts along the modiolar axis, using a vibrating blade. The hemicochlea was bathed in L-15 medium in an experimental chamber, mounted on the stage of an upright microscope (Leica DMLB). A glass paddle was placed below the basilar membrane in the hemi-turn, that was chosen for measurements (Fig. 1a). Glass paddles were pulled from 1.5-mm OD glass tubing to a tip diameter of $50 \mu\text{m}$. The paddle, placed close to the middle point of the pectinate zone of the basilar membrane in scala tympani, was vibrated at 102 Hz . The mechanical stimulus was coupled through the fluids to the basilar membrane. Quality control of hemicochleae was based on the physical appearance of the preparation. Specifically, we consider a preparation 'good' if the pillar cell phalanges are straight, the OHCs are cylindrical and straight, there is no bending of the basilar membrane between its pectinate and arcuate zones, and most importantly, the tectorial membrane is closely apposed to the third row of OHCs. Another marker for the hemicochlea's condition is the shape of the OHCs. Because OHCs became significantly swollen in less than 25 min after the hemicochlea is exposed to the extracellular medium, all the data presented were collected within 15 min. All experiments were performed at room temperature ($22 \pm 2^\circ\text{C}$).

The hemicochlea was bathed in the modified L-15 medium, which contained: 136 mM NaCl , $5.8 \text{ mM NaH}_2\text{PO}_4$, 5.4 mM KCl , 1.6 mM CaCl_2 , 0.9 mM MgCl_2 , 0.4 mM MgSO_4 , 10 mM HEPES-NaOH at pH 7.4 and $300 \text{ mosmol l}^{-1}$. The pipette solution contained: 140 mM CsCl , 0.1 mM CaCl_2 , 3.5 mM MgCl_2 , 2.5 mM MgATP , 5 mM EGTA-KOH , 5 mM HEPES-KOH at pH 7.4 and $300 \text{ mosmol l}^{-1}$. When normal intracellular medium was used to record receptor current or receptor potential, 140 mM KCl was used to replace an equal amount of CsCl in the patch pipette. The resting membrane potential of OHCs, on rupturing, was $27 \pm 9 \text{ mV}$ ($n = 56$), which rapidly hyperpolarized to $-57 \pm 12 \text{ mV}$ after equilibration with the content of the patch pipette. In some recordings, when the influence of gating current from the motor prestin was to be eliminated, 140 mM pentane-sulfonate was used to replace the equal amount of Cl^- in the patch electrode^{21,31}. Pipette resistances were 3 – $4 \text{ M}\Omega$. Recordings were made in whole-cell voltage-clamp mode, using an Axopatch 200B patch-clamp amplifier (Axon Instruments). Series resistance was 8 – $10 \text{ M}\Omega$, and approximately 75% of series resistance was compensated. Currents were filtered at 1 kHz , and digitized at 5 kHz with a 16-bit A/D converter (Digidata 1322B) and pClamp 9.01 software (Axon Instruments). The cells were held at -70 mV . Basilar membrane motion was evoked by a series of 102 Hz sinusoidal bursts with different

amplitudes. Basilar membrane displacement was measured by a photodiode system³¹ mounted on the photo port of the microscope. In brief, the magnified image of the edge of the basilar membrane was projected onto the photodiode through a rectangular slit. The photocurrent response was calibrated to displacement units by moving the slit by a fixed distance ($0.5 \mu\text{m}$) with the image in the slit at the beginning of each trial. The photocurrent signal was low-pass filtered (set at $1,000 \text{ Hz}$) by an anti-aliasing filter before being digitized by a 16-bit A/D converter (Digidata 1322B). The photodiode system itself had a cutoff (-3 dB) frequency of $1,100 \text{ Hz}$. With averaging five samples and low-pass filtering, movement magnitudes as low as 15 nm could be detected.

Received 9 January; accepted 23 April 2004; doi:10.1038/nature02591.

- Dallos, P. The active cochlea. *J. Neurosci.* **12**, 4575–4585 (1992).
- Brownell, W. E., Bader, C. R., Bertrand, D. & de Ribaupierre, Y. Evoked mechanical responses in isolated cochlear outer hair cells. *Science* **227**, 194–196 (1985).
- Lieberman, M. C. *et al.* Prestin is required for outer hair cell electromotility and the cochlear amplifier. *Nature* **419**, 300–304 (2002).
- Hudspeth, A. J. & Corey, D. P. Sensitivity, polarity and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli. *Proc. Natl Acad. Sci. USA* **74**, 2407–2411 (1977).
- Denk, W., Holt, J. R., Shepherd, G. M. & Corey, D. P. Calcium imaging of single stereocilia in hair cells: localization of transduction channels at both ends of tip links. *Neuron* **15**, 1311–1321 (1995).
- Richter, C. P., Evans, B. N., Edge, R. & Dallos, P. Basilar membrane vibration in the gerbil hemicochlea. *J. Neurophysiol.* **79**, 2255–2264 (1998).
- Ashmore, J. F., Kolston, P. J. & Mammano, F. In *Biophysics of Hair Cell Sensory Systems* (eds Duifhuis, H., Horst, J. W., Dijk, P. & van Netten, S. M.) 151–158 (World Scientific, Singapore, 1993).
- Kros, C. J., Rüsch, A. & Richardson, G. P. Mechano-electrical transducer currents in hair cells of the cultured neonatal mouse cochlea. *Proc. R. Soc. Lond. B* **249**, 185–193 (1992).
- Kennedy, H. J., Evans, M. G., Crawford, A. C. & Fettiplace, R. Fast adaptation of mechano-electrical transducer channels in mammalian cochlear hair cells. *Nature Neurosci.* **6**, 832–836 (2003).
- Müller, M. The cochlear place-frequency map of the adult and developing Mongolian gerbil. *Hear. Res.* **94**, 148–156 (1996).
- Ricci, A. J., Crawford, A. C. & Fettiplace, R. Tonotopic variation in the conductance of the hair cell mechanotransducer channel. *Neuron* **40**, 983–990 (2003).
- Russell, I. J. & Sellick, P. M. Low-frequency characteristics of intracellularly recorded receptor potentials in guinea-pig cochlear hair cells. *J. Physiol. (Lond.)* **338**, 179–206 (1983).
- Russell, I. J., Cody, A. R. & Richardson, G. P. The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown *in vitro*. *Hear. Res.* **22**, 199–216 (1986).
- Dallos, P. Response characteristics of mammalian cochlear hair cells. *J. Neurosci.* **5**, 1591–1608 (1985).
- Dallos, P. Organ of Corti kinematics. *J. Assoc. Res. Otolaryngol.* **4**, 416–421 (2003).
- Géléoc, G. S. *et al.* A quantitative comparison of mechano-electrical transduction in vestibular and auditory hair cells of neonatal mice. *Proc. R. Soc. Lond. B* **264**, 611–621 (1997).
- Cheatham, M. A. *et al.* Assessment of hair-cell transducer function in prestin knockout mice. *Assoc. Res. Otolaryngol. Abstr.* **1248** (2004).
- Gale, J. E. & Ashmore, J. F. Charge displacement induced by rapid stretch in the basolateral membrane of the guinea-pig outer hair cell. *Proc. R. Soc. Lond. B* **255**, 243–249 (1994).
- Zheng, J. *et al.* Prestin is the motor protein of cochlear outer hair cells. *Nature* **405**, 149–155 (2000).
- Corey, D. P. & Hudspeth, A. J. Ionic basis of the receptor potential in a vertebrate hair cell. *Nature* **281**, 675–677 (1979).
- Oliver, D. *et al.* Intracellular anions as the voltage sensor of prestin, the outer hair cell motor protein. *Science* **292**, 2340–2343 (2001).
- Crawford, A. C., Evans, M. G. & Fettiplace, R. The actions of calcium on the mechano-electrical transducer current of turtle hair cells. *J. Physiol. (Lond.)* **434**, 369–398 (1991).
- Dallos, P., Santos-Sacchi, J. & Flock, Å. Intracellular recordings from cochlear outer hair cells. *Science* **218**, 582–584 (1982).
- Russell, I. J., Richardson, G. P. & Cody, A. R. Mechanosensitivity of mammalian auditory hair cells *in vitro*. *Nature* **321**, 517–519 (1986).
- Housley, G. D. & Ashmore, J. F. Ionic currents of outer hair cells isolated from the guinea-pig cochlea. *J. Physiol. (Lond.)* **448**, 73–98 (1992).
- Crawford, A. C. & Fettiplace, R. An electrical tuning mechanism in turtle cochlear hair cells. *J. Physiol. (Lond.)* **312**, 377–412 (1981).
- Holt, J. R. *et al.* A chemical-genetic strategy implicates myosin-1c in adaptation by hair cells. *Cell* **108**, 371–381 (2002).
- Mammano, F. & Ashmore, J. F. Differential expression of outer hair cell potassium currents in the isolated cochlea of the guinea-pig. *J. Physiol. (Lond.)* **496**, 639–646 (1996).
- Dallos, P. Neurobiology of cochlear inner and outer hair cells: intracellular recordings. *Hear. Res.* **22**, 185–198 (1986).
- Cooper, N. P. & Rhode, W. S. Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea. *J. Neurophysiol.* **78**, 261–270 (1997).
- He, D. Z. Z., Jia, S. P. & Dallos, P. Prestin and the dynamic stiffness of cochlear outer hair cells. *J. Neurosci.* **23**, 9089–9096 (2003).

Acknowledgements This work has been supported by National Institute of Health grants to D.Z.Z.H., and to P.D.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.Z.Z.H. (hed@creighton.edu).

Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR- γ

Frédéric Picard¹, Martin Kurtev¹, Namjin Chung¹, Acharawan Topark-Ngarm², Thanaset Senawong², Rita Machado de Oliveira^{1,3}, Mark Leid², Michael W. McBurney⁴ & Leonard Guarente¹

¹Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

²Laboratory of Molecular Pharmacology, Department of Pharmaceutical Sciences, College of Pharmacy and Environmental Health Science Center, Oregon State University, Corvallis, Oregon 97331-3507, USA

³Graduate Program in Basic and Applied Biology, ICBAS, University of Porto, 4099-003, Portugal

⁴Ottawa Regional Cancer Centre and Department of Medicine, University of Ottawa, Ontario K1H 1C4, Canada

Calorie restriction extends lifespan in organisms ranging from yeast to mammals¹. In yeast, the *SIR2* gene mediates the life-extending effects of calorie restriction². Here we show that the mammalian *SIR2* orthologue, *Sirt1* (sirtuin 1), activates a critical component of calorie restriction in mammals; that is, fat mobilization in white adipocytes. Upon food withdrawal *Sirt1* protein binds to and represses genes controlled by the fat regulator PPAR- γ (peroxisome proliferator-activated receptor- γ), including genes mediating fat storage. *Sirt1* represses PPAR- γ by docking with its cofactors NCoR (nuclear receptor co-repressor) and SMRT (silencing mediator of retinoid and thyroid hormone receptors). Mobilization of fatty acids from white adipocytes upon fasting is compromised in *Sirt1*^{+/-} mice. Repression of PPAR- γ by *Sirt1* is also evident in 3T3-L1 adipocytes, where overexpression of *Sirt1* attenuates adipogenesis, and RNA interference of *Sirt1* enhances it. In differentiated fat cells, upregulation of *Sirt1* triggers lipolysis and loss of fat. As a reduction in fat is sufficient to extend murine lifespan³, our results provide a possible molecular pathway connecting calorie restriction to life extension in mammals.

Calorie restriction is the first regimen shown to extend mammalian lifespan. Calorie restriction induces a shedding of body fat from white adipose tissue (WAT), a decrease in body temperature and an increase in insulin sensitivity⁴. It has been suggested that calorie restriction functions in a passive way by lowering oxidative and other damage, thereby resulting in a longer lifespan. However, studies in yeast suggest that calorie restriction is a highly regulated response that requires a sensing step followed by the execution of a programme to extend lifespan⁵. The regulatory gene that mediates this programme is *SIR2*, which promotes survival in yeast⁶, *Caenorhabditis elegans*⁷ and, perhaps, higher eukaryotes in response to food scarcity. *SIR2* is well suited for this function because it encodes an NAD-dependent protein deacetylase¹, enabling it to monitor cellular metabolism and exert corresponding effects on gene expression.

WAT seems to be a primary factor in longevity, as mice engineered to have reduced levels of it live longer³. This effect on lifespan may result from changes in hormones that are normally secreted from WAT in proportion to fat mass. Thus, we evaluated whether the mammalian *Sir2* orthologue, *Sirt1*, senses nutrient availability in WAT and mediates corresponding effects on fat accumulation. To probe whether *Sirt1* may actually modulate adipogenesis, we used mouse 3T3-L1 fibroblasts as an *in vitro* model⁸. Adipogenesis in these cells is promoted by the nuclear receptor PPAR- γ ⁹. Upon induction of adipogenesis by insulin, dexamethasone and isobutyl-

methylxanthine, we observed that *Sirt1* protein levels increased and peaked at day 5 after hormonal stimulation (Fig. 1a). *Sirt1* expression in 3T3-L1 cells was then modified through retroviral infection with either pBABE-*Sirt1* or pSUPER-*Sirt1* RNA interference (RNAi) for overexpression (tenfold) or downregulation (sevenfold) of the *Sirt1* gene, respectively (Fig. 1b). 3T3-L1 cells undergo one or two mitotic divisions after induction as a prelude to terminal differentiation^{9,10}. This occurred normally in cells that overexpressed or underexpressed *Sirt1* as evaluated by 5-bromo-deoxyuridine (BrdU) incorporation (data not shown). However, compared with cells infected with the control vector, stable 3T3-L1 cells overexpressing *Sirt1* accumulated much less fat as determined by Oil red O staining after 7 days of differentiation (Fig. 1c) or direct measurement of intracellular triglyceride content (Fig. 1d). In contrast, downregulation of *Sirt1* expression resulted in a significant increase in triglyceride accumulation after differentiation (Fig. 1c, d). These results indicate that *Sirt1* acts a negative modulator of adipogenesis in the 3T3-L1 model.

Insulin/insulin-like growth factor (IGF)-1 signalling helps to regulate adipogenesis¹¹, and this pathway determines lifespan in nematode worms¹². To evaluate the possible interaction of *Sirt1* with the insulin/IGF-1 signalling cascade in adipocytes, we differentiated virally transduced 3T3-L1 cells in the absence of insulin but with rosiglitazone, a very potent, selective PPAR- γ agonist that acts downstream of the insulin pathway. Although rosiglitazone treatment promoted adipogenesis to a greater extent than the regular differentiation cocktail (Fig. 1c), it did not alter the phenotypes of cells with increased or decreased levels of *Sirt1* (Fig. 1c), suggesting that *Sirt1* affects regulators that act downstream of insulin/IGF-1 signalling.

Differentiation of 3T3-L1 fibroblasts increases levels of the transcription factor C/EBP- δ , which stimulates the expression of PPAR- γ and C/EBP- α ¹⁰. PPAR- γ induces expression of target genes, such as the fatty-acid-binding protein aP2 (also known as adipose tissue-specific FABP)¹³. Moreover, PPAR- γ can maintain expression of itself, perhaps by binding to PPAR- γ sites in the promoter of the PPAR- γ gene (*Pparg*)^{9,14}. To gain an insight into the mechanisms by which *Sirt1* represses fat accretion, we measured protein and messenger RNA expression of key factors in the transcriptional programme in the different virus-infected 3T3-L1 adipocytes. We observed a reduction in C/EBP- α , C/EBP- δ and aP2 mRNA, but not C/EBP- β , upon *Sirt1* overexpression (Fig. 1e). A reduction in PPAR- γ and C/EBP- α was also observed by western blotting (Fig. 1f). In contrast, cells in which *Sirt1* had been downregulated showed higher levels of PPAR- γ , C/EBP- δ , C/EBP- α and aP2 (Fig. 1e, f). These findings are consistent with the model that *Sirt1* functions as a repressor of genes that drive white adipocyte differentiation and fat storage.

The above experiments do not address whether *Sirt1* regulates fat accumulation in fully differentiated adipocytes. Thus, we fully differentiated 3T3-L1 cells and subsequently (12 days after induction) applied the *Sirt1* activator resveratrol¹⁵ over a wide range of concentrations (Fig. 2a). After staining the cells for fat content, a strong reduction in fat was observed at 50 and 100 μ M resveratrol. The loss of fat was due to activation of *Sirt1*, because there was no drug-mediated fat reduction in cells in which *Sirt1* levels were knocked down (Fig. 2b). To validate further these visual results, we measured triglyceride content and free fatty acid (FFA) release in these cells. Triglyceride content was reduced and release of FFA was stimulated by resveratrol in the control cells, but not in the *Sirt1* knockdown cells (Fig. 2c). These findings strongly suggest that upregulation of *Sirt1* stimulates fat mobilization in fully differentiated 3T3-L1 adipocytes.

To determine whether *Sirt1* stimulates fat mobilization in bona fide adipocytes, we cultured primary rat white adipocytes and activated fat mobilization with the known β -adrenergic inducer adrenalin. Addition of resveratrol greatly stimulated the release of FFA triggered by adrenalin (Fig. 2d), consistent with the findings in

the 3T3-L1 cells. In a converse experiment using these rat adipocytes, we addressed whether inhibition of Sirt1 would blunt the mobilization of fat by adrenalin. Addition of the known Sirt1 inhibitor nicotinamide¹⁶ indeed reduced the release of FFA triggered by adrenalin (Fig. 2e). The above findings indicate that Sirt1 not only represses the differentiation programme of adipocytes, but also activates the mobilization of fat in fully differentiated cells.

Sirt1 is an NAD-dependent deacetylase that can repress activity of p53 (ref. 1) and forkhead proteins^{17,18}. Repression of adipogenesis and fat retention in 3T3-L1 cells by Sirt1 might be explained by inhibition of another transcription factor, PPAR- γ , as its activity is crucial for differentiation and maintenance of adipocytes^{9,14,19}. Thus, we determined by chromatin immunoprecipitation (ChIP)

assays whether Sirt1 binds to PPAR- γ sites in the promoters of the *aP2* and *Pparg* genes. In 3T3-L1 cells, Sirt1 and PPAR- γ were both bound to similar promoter regions of *aP2* and *Pparg* (Fig. 3a). The interaction of Sirt1 with both promoter sequences was stronger in Sirt1-overexpressing cells and was lost in Sirt1 knockdown cells (Fig. 3a). Binding of Sirt1 to a control DNA region 2.5 kilobases (kb) upstream of the PPAR- γ site near *Pparg* was not observed (data not shown). Furthermore, luciferase reporter assays showed that Sirt1 repressed transactivation by PPAR- γ (Fig. 3b). These findings indicate that Sirt1 and PPAR- γ bind to the same DNA sequences and suggest that Sirt1 is a co-repressor of PPAR- γ .

To gain further evidence that Sirt1 is a PPAR- γ co-repressor, we next determined whether the two proteins interact. Sirt1 was

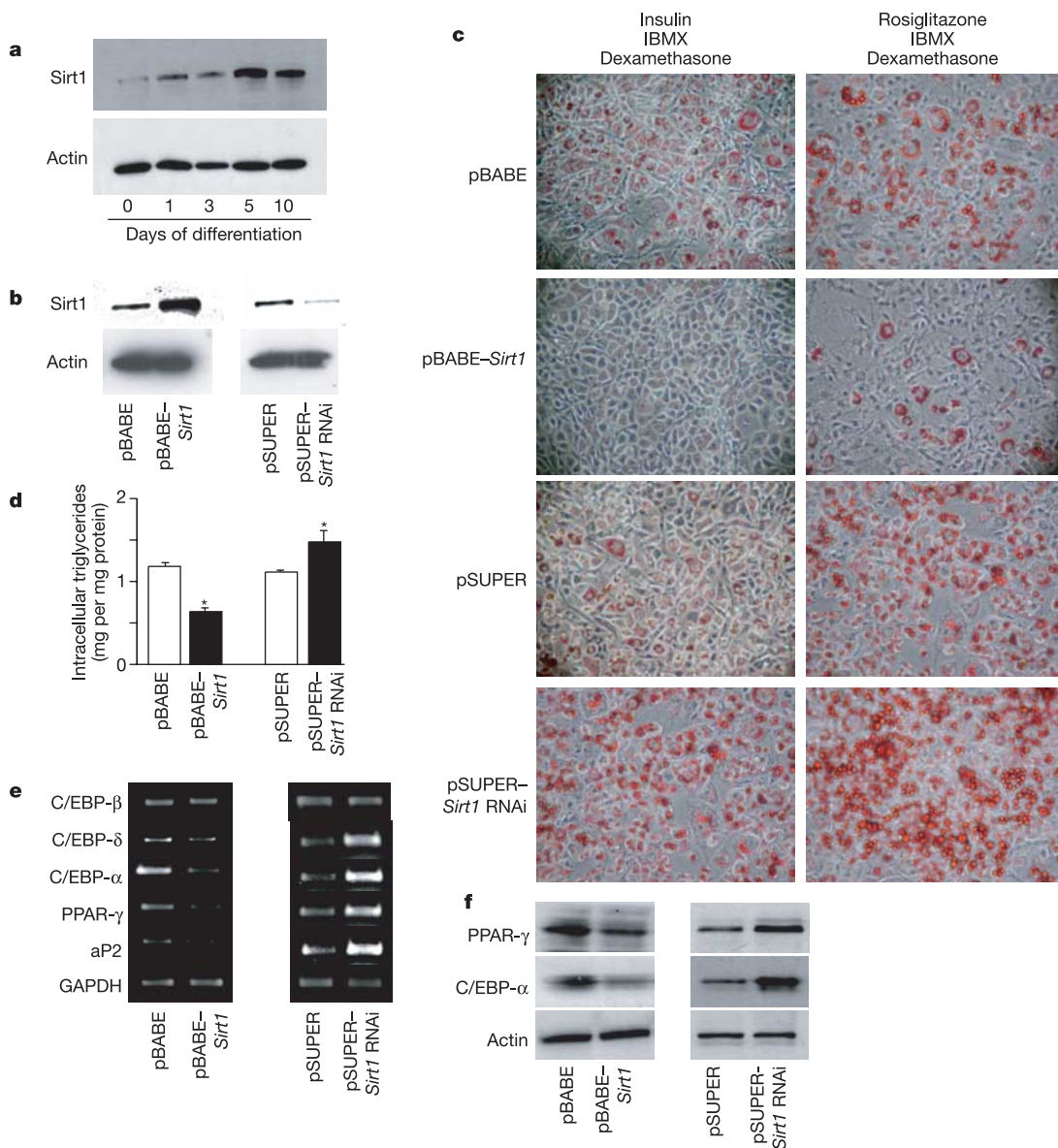


Figure 1 Sirt1 regulates adipogenesis and triglyceride accumulation in 3T3-L1 cells. **a**, Time course of Sirt1 protein expression on a western blot in 3T3-L1 cells upon induction of adipogenesis. **b**, Sirt1 protein levels in 3T3-L1 cells infected with control pBABE or pSUPER retroviral vectors or with pBABE-Sirt1 or pSUPER-Sirt1 RNAi vectors for overexpression or downregulation of Sirt1, respectively. **c**, Oil red O staining for triglycerides in infected 3T3-L1 cells differentiated for 7 days with IBMX and dexamethasone plus either insulin or rosiglitazone. **d**, Intracellular triglyceride levels in infected 3T3-L1 cells differentiated for 7 days with insulin, IBMX and dexamethasone.

Data shown are means + s.e.m. and were obtained from two independent experiments carried out in triplicate. Asterisk indicates a statistical difference from respective control ($P < 0.05$). **e**, **f**, 3T3-L1 cells were infected and differentiated as in **d**. Protein and total RNA were extracted at day 7. **e**, Reverse transcription and PCRs were on complementary DNAs with gene-specific primers. One microgram of total RNA was used in each case. **f**, Protein levels of PPAR- γ and C/EBP- α in total cell extracts measured by western blotting. GAPDH and actin levels are shown as loading controls.

detected in PPAR- γ immunoprecipitates from differentiated 3T3-L1 adipocytes (Fig. 3c). The PPAR- γ cofactor NCoR²⁰ was also co-immunoprecipitated by anti-Sirt1 antiserum but not by pre-immune serum (Fig. 3d). Glutathione S-transferase (GST) pull-down experiments revealed that two NCoR fragments interact with Sirt1: repression domain 1 (RD1) and the CBF/Su(H) interaction domain (Fig. 3e). Reciprocally, NCoR RD1 was found to interact with the amino-terminal region of Sirt1 (amino acids 1–214, GST-Sirt1(Nt); Fig. 3f), and the CBF1/Su(H) interaction domain of NCoR interacted with the homology domain of Sirt1 (amino acids 214–541, GST-Sirt1(SHD); Fig. 3f). In analogous experiments, GST pull-down assays revealed an interaction between the Sirt1 homology domain and SMRT (Fig. 3g).

The interaction between Sirt1 and NCoR (and SMRT) suggests that Sirt1 represses PPAR- γ activity by docking with the cofactors. Consistent with this, ChIP assays revealed that NCoR also binds to known PPAR- γ sites of promoters of adipogenic genes (Fig. 4h). To test the possibility that Sirt1 functionally represses PPAR- γ by means of NCoR, we infected Sirt1-overexpressing 3T3-L1 fibroblasts with an NCoR RNAi virus. Western blots showed that doubly infected cells overexpressed Sirt1 and underexpressed NCoR (not shown). As expected, overexpression of Sirt1 in the absence of the NCoR RNAi virus reduced fat accretion (Fig. 4i). In contrast, this reduction was largely prevented on simultaneous treatment with NCoR RNAi (Fig. 4i), demonstrating that NCoR is required for repression of fat accumulation by Sirt1.

To test for a role of Sirt1 in fat mobilization *in vivo*, we first determined whether Sirt1 was expressed in WAT in mice. Sirt1 was found in all white adipose depots examined, with no notable distribution differences (Fig. 4a). Next, to probe for an *in vivo* function, we used mice with germline mutations in the *Sirt1* gene. Unfortunately, the total absence of Sirt1 in mice (*Sirt1*^{-/-}) results in a high degree of post-natal lethality and other severe phenotypes, precluding their use in this study²¹. Therefore, we compared wild type to *Sirt1*^{+/-} mice, which are phenotypically normal. No significant differences in epididymal WAT mass were observed between the wild type and heterozygous cohorts (Fig. 4b). As yeast Sir2 is important during calorie restriction, we assayed by ChIP the recruitment of Sirt1 to PPAR- γ -binding sites in the aP2 and PPAR- γ promoters in WAT of mice that were either fed or fasted. In mice fed *ad libitum*, Sirt1 was not bound to aP2 or PPAR- γ promoter sequences (Fig. 4c). However, Sirt1 was bound to these sequences after overnight food deprivation, showing that Sirt1 is recruited to PPAR- γ DNA-binding sites upon fasting.

As a test of the effect of Sirt1 on fatty acid mobilization in adipocytes, we next addressed whether fatty acid release from WAT upon fasting was altered in *Sirt1*^{+/-} mice. Heterozygous gene ablation was associated with a 40–45% lowering in circulating FFA levels in the blood after overnight food deprivation compared with wild type (Fig. 4d, $P < 0.05$). To verify that these effects of *Sirt1* genotype were due to differences in fat release from WAT and not reuptake of FFA from the blood by oxidative tissues, we cultured the

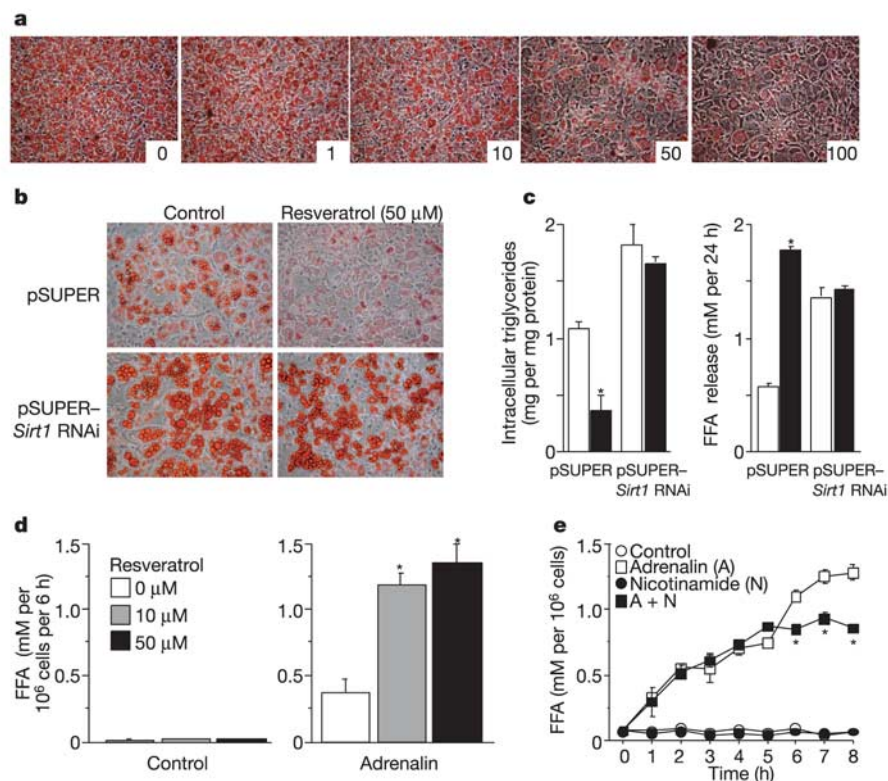


Figure 2 Activation of Sirt1 by resveratrol decreases fat accumulation in differentiated adipocytes. **a**, Resveratrol addition to 3T3-L1 cells first fully differentiated into adipocytes. At day 12, resveratrol dissolved in DMSO was added in increasing concentrations (1, 10, 50 and 100 μ M; as indicated) and Oil red O staining was performed 3 days later and compared with cells incubated with DMSO alone (resveratrol 0 μ M). **b**, **c**, Effect of resveratrol on 3T3-L1 cells infected with the indicated viruses. Cells were differentiated into adipocytes as in Fig. 1d. At day 12, resveratrol (50 μ M) was added and 3 days later, intracellular triglyceride content was evaluated by Oil red O staining (**b**) and in

cell lysates (**c**). FFA released in the medium during the 3-day incubation period was also measured (**c**). Open bars, control; filled bars, resveratrol. **d**, **e**, Effect of resveratrol on primary adipocytes isolated from rat retroperitoneal WAT. Cells were incubated with or without adrenalin (10^{-7} and 10^{-5} M for **d** and **e**, respectively) in the presence of Sirt1 activator (resveratrol, **d**) or inhibitor (nicotinamide, **e**). FFA levels released in the medium were measured after 4 h (**d**) or at indicated time points (**e**). Data shown are means $\pm$ s.e.m. and were obtained from two independent experiments done in triplicate. Asterisk indicates a statistical difference from respective control ($P < 0.05$).

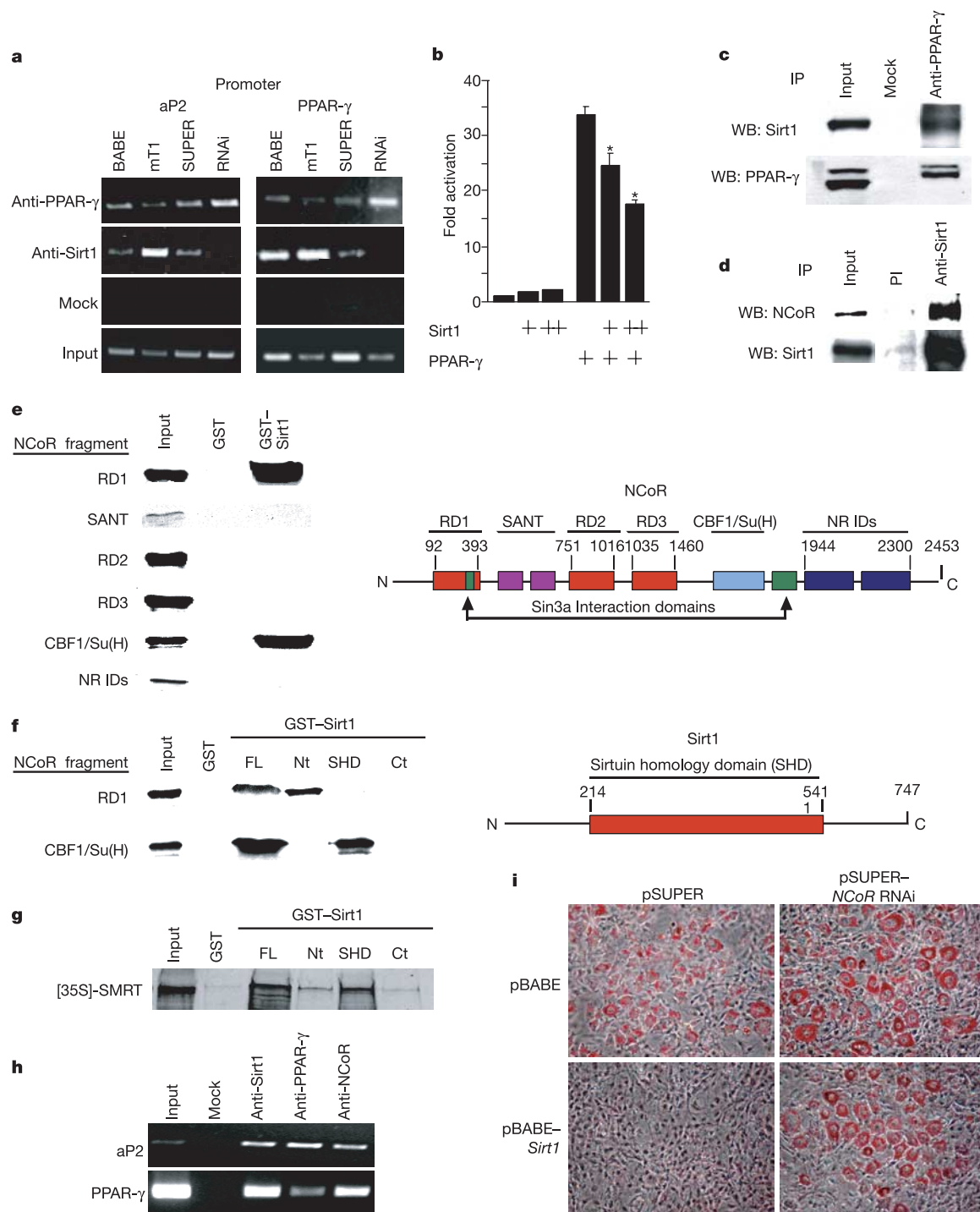


Figure 3 Sirt1 reduces fat accumulation by repressing PPAR- γ activity through Sirt1–NCoR interactions. **a**, ChIP assays on 3T3-L1 cells infected as in Fig. 1c. Cells were differentiated for 7 days and assayed (Methods). **b**, Luciferase assays of 293T cells transfected with mouse PPAR- γ 2 and pGL3-PPRE and with increasing concentrations of mouse *Sirt1*. Luciferase was measured 24 h later and was corrected for transfection efficiency. Data represent mean $\pm$ s.e.m. of triplicate experiments carried out in duplicate. Asterisk indicates a statistical difference ($P < 0.05$). **c**, Co-immunoprecipitation of endogenous PPAR- γ and Sirt1 proteins in whole-cell extracts prepared from uninfected differentiated 3T3-L1 adipocytes. **d**, Co-immunoprecipitation of endogenous Sirt1 and NCoR proteins in HEK293 whole-cell extracts using pre-immune (PI) or anti-Sirt1 serum. **e**, Mapping the Sirt1 interaction interface of NCoR. Constructs encoding the indicated NCoR fragments were 35 S-methionine-translated and used for GST pull-down experiments as previously

described³⁰. A schematic representation of NCoR functional domains is shown on the right. NR IDs, nuclear receptor interaction domains. SANT; Swi3, Ada2, NCoR, TF11B shared motif. **f**, Mapping the NCoR interaction interface of Sirt1. GST pull-down experiments were conducted using radiolabelled NCoR RD1 and CBF1/Su(H) interaction domains and the indicated fragments of Sirt1 fused to GST. A schematic representation of full-length Sirt1 is shown on the right. FL, full length; Nt, N-terminal; Ct, C-terminal. **g**, GST pull-down between the indicated fragments of Sirt1 fused to GST and 35 S-methionine-translated SMRT. **h**, ChIP assay in uninfected 3T3-L1 cells differentiated for 7 days with insulin, IBMX and dexamethasone. **i**, RNAi of NCoR. 3T3-L1 cells were infected with either pBABE or pBABE-*Sirt1* and co-infected with either pSUPER or pSUPER-*NCoR* RNAi. Puromycin-selected cells were then differentiated with insulin, IBMX and dexamethasone and stained with Oil red O 7 days later.

same number of white adipocytes from *Sirt1*^{+/+} and *Sirt1*^{+/-} mice, challenged them with adrenalin, and measured the release of FFA. Again, FFA release was reduced in the *Sirt1*^{+/-} cells compared with the *Sirt1*^{+/+} cells (Fig. 4e).

Here we show that the mammalian Sir2 orthologue, Sirt1, is activated by food deprivation to trigger fat mobilization in WAT. The pharmacological activation of Sirt1 also elicits the lipolysis of triglycerides and the release of FFA. Because a reduction in fat storage in WAT is a primary way by which calorie restriction extends lifespan in mammals, our results provide a possible mechanism for understanding the regulation of mammalian lifespan by diet. Sirt1 represses WAT by inhibiting the nuclear receptor PPAR- γ ^{13,19}.

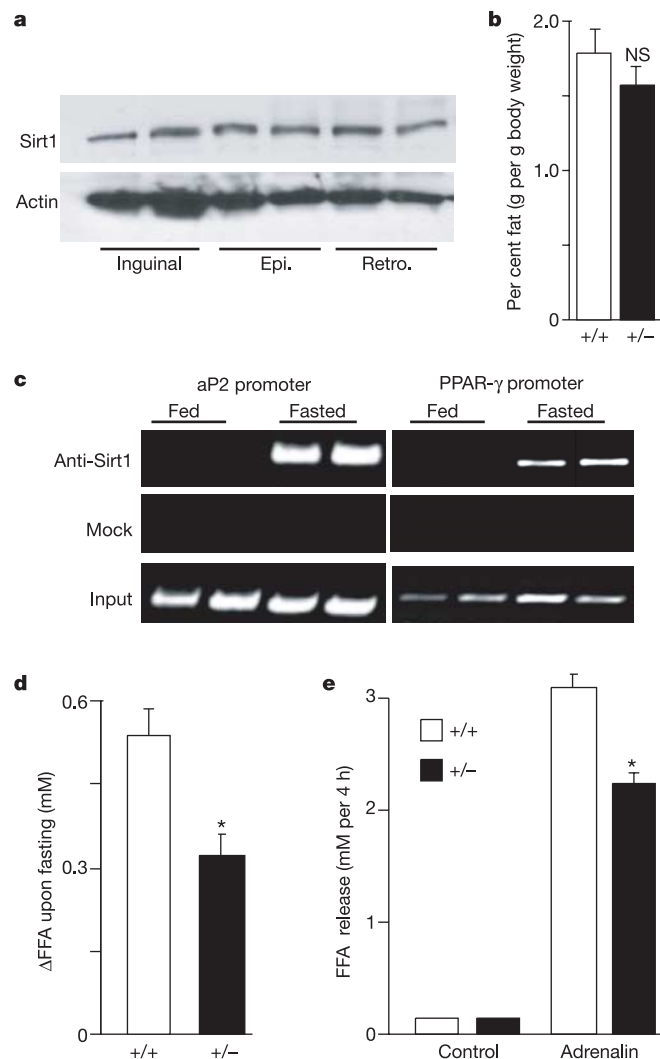


Figure 4 Sirt1 promotes fat mobilization *in vivo*. **a**, Protein levels of Sirt1 in subcutaneous inguinal and visceral epididymal (Epi.) and retroperitoneal (Retro.) WAT depots in wild-type fed mice. Actin levels are shown as loading controls. Each lane represents an individual mouse. **b**, Weight of epididymal WAT depots in fed *Sirt1*^{+/+} and *Sirt1*^{+/-} mice. Bars represent mean $\pm$ s.e.m. ($n = 10$ – 12 ; NS, not significant). **c**, *In vivo* ChIP assay in FVB mice either fed or fasted overnight. Each lane represents an individual mouse. Findings represent five independent experiments done in duplicate. **d**, Change in FFA levels in blood from age-matched male *Sirt1*^{+/+} and *Sirt1*^{+/-} mice between before and after overnight fasting. Bars represent mean $\pm$ s.e.m. of 27–30 animals. Asterisk indicates a statistical difference from wild-type littermates ($P < 0.05$). **e**, Isolated adipocytes from *Sirt1*^{+/+} and *Sirt1*^{+/-} mice incubated with or without adrenalin (10^{-5} M). After 4 h of incubation, FFA released in the medium were measured. Bars represent mean $\pm$ s.e.m. of triplicate experiments. Asterisk indicates a statistical difference from *Sirt1*^{+/+} adipocytes ($P < 0.05$).

Starvation of animals causes Sirt1 to interact with PPAR- γ DNA-binding sites and thereby repress target genes that drive fat storage¹⁹. We do not yet know whether Sirt1 is activated upon fasting by a change in the NAD/NADH ratio, as has been proposed in yeast²², or by other metabolic changes. It is also uncertain whether Sirt1 deacetylates PPAR- γ , histones at target genes, or both.

The pathway of regulation described here may impact on age-related diseases. The accumulation of WAT during ageing is associated with several adverse complications, such as insulin resistance, type 2 diabetes and atherosclerosis²³. Given the impact of Sirt1 on PPAR- γ activity and because PPAR- γ activity helps determine age-related insulin resistance²⁴, Sirt1 may have an important role in metabolic diseases and link the effects of food consumption to body fat mass and diseases of ageing. It is likely that calorie restriction exerts other effects on mammals to increase longevity, besides reducing WAT, as longevity in mice with reduced fat is not as great as animals on a long-term calorie restriction regimen. Tissues that metabolize fat and carbohydrate may also be important in delivering some of the benefit of calorie restriction, and it will be of interest to determine whether Sirt1 upregulates metabolism upon food reduction to round out an optimal profile for long life. $\square$

Methods

Animal experimentation

Wild-type FVB male age-matched (12–16 weeks old) mice were used for the present studies. *Sirt1*^{+/+} and *Sirt1*^{+/-} genotypes have been described previously²¹. All mice were housed under controlled temperature ($25 \pm 1^\circ\text{C}$ (s.d.)) and lighting conditions. Food provided was normal chow. All mice were cared for in accordance with the MIT animal care committee. Blood was collected from the retro-orbital sinus and kept on ice until centrifugation (1,500 g, 15 min at 4°C), and the plasma was stored at -20°C until analysis. All animals were killed by decapitation. WAT depots were collected, weighed and quickly frozen in liquid nitrogen and stored at -70°C until further processing by immunoblotting. Non-esterified free fatty acids were determined by enzymatic assays (Wako Pure Chemical Industries).

Cell culture, retroviral infection and transfection

3T3-L1, HEK293, 293T and Phoenix cells (ATCC, Rockville, MD) were cultured in Dulbecco's modified Eagle's medium with 10% FCS, and antibiotics. Primary adipocytes from Sprague-Dawley rats were prepared as described previously²⁵. Transfections for luciferase assays were done as described^{9,13} using pSPORT6-PPAR- γ 2, pGL3-PPRE²⁶ and pcDNA3-Sirt1. Data were corrected for transfection efficiency.

Retroviral infection was performed as described⁹. Phoenix cells were transfected with either pBABE, pBABE-Sirt1, pSUPERretro (Oligoengine), pSUPERretro-Sirt1 RNAi (5'-GATGAAGTTGACCTCCTCA-3') or pSUPER-NcoR RNAi (5'-GCTGCATCCAAGGGCCATG-3') using Lipofectamine (Invitrogen). After 48 h of transfection, the medium containing retroviruses was collected, filtered, treated by polybrene ($1 \mu\text{g ml}^{-1}$) and transferred to 3T3-L1 target cells. Infected cells were selected with puromycin ($2.5 \mu\text{g ml}^{-1}$) for 7 days.

To stimulate adipogenesis and accumulation of lipids in cells, medium was supplemented to confluent cells (day 0) with $2 \mu\text{M}$ insulin or 10^{-7} M rosiglitazone (Alexis Biochemicals), as stated in the text, $1 \mu\text{M}$ dexamethasone, and 0.25 mM isobutylmethylxanthine (IBMX) for 2 days. The cells were then incubated with insulin or 10^{-7} M rosiglitazone, changing the medium every second day. Fat accumulation was visualized by staining of lipids with Oil red O. Intracellular triglyceride levels were measured in cell lysates by an enzymatic method using a reagent kit from Boehringer Mannheim.

RNA and protein preparation and analysis

Total RNA from cultured cells were extracted (Qiagen) and analysed by semi-quantitative polymerase chain reaction with reverse transcription. GAPDH and 18S RNA levels were determined as a control for loading. Primer sequences have been described elsewhere^{19,27}. Proteins from mouse tissues were extracted in a solution of pH 7.4 containing 20 mM HEPES, 250 mM sucrose, 4 mM EDTA, 1% Triton and protease inhibitor cocktail. 3T3-L1 cells were lysed in NET-N buffer (20 mM Tris-HCl, pH 8 containing 150 mM NaCl, 0.5% NP-40, 10% glycerol, 1 mM EDTA and a protease inhibitor cocktail).

Chromatin immunoprecipitation

For *in vitro* ChIP, infected 3T3-L1 cells were differentiated as mentioned above. For *in vivo* ChIP, epididymal WAT was dissected, minced and fixed overnight in PBS containing 1% formaldehyde and protease inhibitor cocktail. Tissues were then rinsed five times in PBS. Further ChIP assays were performed as described previously²⁸, using 1:200 antibody dilutions to immunoprecipitate DNA-protein complexes. DNA was then purified using Qiagen PCR purification kit and PCR reaction was performed using primers for aP2 (5'-AAATTCAGAAAGTAACACATTATT-3'; 5'-ATGCCTGACCATGTGA-3') and PPAR- γ proximal (amplifying a region at 0.3 kb upstream of ATG: 5'-GAGCAAGGTCTTCATCATTACG-3'; 5'-CCCCTGGAGCTGGAGTTAC-3') and distal

(amplifying a region at 2.8 kb upstream of ATG: 5'-CTCTCCACCTCGGCATAC-3'; 5'-TTGCCAGAGAGCCAGTGACA-3') promoters.

Pull-down and co-immunoprecipitation assays

Differentiated 3T3-L1 or HEK293 cells were lysed in NET-N buffer by agitation at 4 °C for 30 min. After a brief sonication, lysates were cleared by centrifugation and immunoprecipitated. Antibodies used were anti-Sirt1 (Upstate), anti-NCoR (Upstate), anti-PPAR- γ (SantaCruz), anti-C/EBP- α (SantaCruz) and anti-Sirt1 antiserum²⁹ or preimmune serum. Immunoprecipitates were then analysed by immunoblotting. Equivalent amounts of GST or GST-Sirt1 fusion proteins were bound to glutathione-Sepharose (Pharmacia) and incubated with ³⁵S-methionine-labelled NCoR or SMRT fragments prepared using TNT transcription-translation system (Promega). The reactions were washed five times with binding buffer (10 mM Na-HEPES containing 10% glycerol, 1 mM EDTA, 1 mM DTT, 150 mM NaCl and 0.05% NP-40) and bound proteins were eluted and resolved on denaturing SDS-polyacrylamide gel electrophoresis gels for analysis by autoradiography.

Statistical analysis

The main and interactive effects were analysed by analysis of variance (ANOVA) factorial or repeated measures when appropriate. When justified by the ANOVA analysis, differences between individual group means were analysed by Fisher's PLSD test. Differences were considered statistically significant at $P < 0.05$.

Received 28 January; accepted 16 April 2004; doi:10.1038/nature02583.

Published online 2 June 2004.

- Koubova, J. & Guarente, L. How does calorie restriction work? *Genes Dev.* **17**, 313–321 (2003).
- Lin, S. J., Defossez, P. A. & Guarente, L. Requirement of NAD and SIR2 for life-span extension by calorie restriction in *Saccharomyces cerevisiae*. *Science* **289**, 2126–2128 (2000).
- Blüher, M., Kahn, B. B. & Kahn, C. R. Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science* **299**, 572–574 (2003).
- Weindruch, R. & Walford, R. L. *The Retardation of Aging and Disease by Dietary Restriction* (C. C. Thomas, Springfield, Illinois, 1988).
- Lin, S. J. *et al.* Calorie restriction extends *Saccharomyces cerevisiae* lifespan by increasing respiration. *Nature* **418**, 344–348 (2002).
- Kaeberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).
- Tissenbaum, H. A. & Guarente, L. Increased dosage of a sir-2 gene extends lifespan in *Caenorhabditis elegans*. *Nature* **410**, 227–230 (2001).
- Green, H. & Kehinde, O. Sublines of mouse 3T3 cells that accumulate lipid. *Cell* **1**, 113–116 (1974).
- Tontonoz, P., Hu, E. & Spiegelman, B. M. Stimulation of adipogenesis in fibroblasts by PPAR γ 2, a lipid-activated transcription factor. *Cell* **79**, 1147–1156 (1994).
- Rosen, E. D. & Spiegelman, B. M. Molecular regulation of adipogenesis. *Annu. Rev. Cell Dev. Biol.* **16**, 145–171 (2000).
- Nakae, J. *et al.* The forkhead transcription factor Foxo1 regulates adipocyte differentiation. *Dev. Cell* **4**, 119–129 (2003).
- Kenyon, C. A conserved regulatory system for aging. *Cell* **105**, 165–168 (2001).
- Tontonoz, P., Hu, E., Graves, R. A., Budavari, A. I. & Spiegelman, B. M. mPPAR γ 2: tissue-specific regulator of an adipocyte enhancer. *Genes Dev.* **8**, 1224–1234 (1994).
- Ren, D., Collingwood, T. N., Rebar, E. J., Wolfe, A. P. & Camp, H. S. PPAR γ knockdown by engineered transcription factors: exogenous PPAR γ 2 but not PPAR γ 1 reactivates adipogenesis. *Genes Dev.* **16**, 27–32 (2002).
- Howitz, K. T. *et al.* Small molecule activators of sirtuins extend *Saccharomyces cerevisiae* lifespan. *Nature* **425**, 191–196 (2003).
- Fulco, M. *et al.* Sir2 regulates skeletal muscle differentiation as a potential sensor of the redox state. *Mol. Cell* **12**, 51–62 (2003).
- Motta, M. C. *et al.* Mammalian SIRT1 represses forkhead transcription factors. *Cell* **116**, 551–563 (2004).
- Brunet, A. *et al.* Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science* **303**, 2011–2015 (2004).
- Tamori, Y., Masugi, J., Nishino, N. & Kasuga, M. Role of peroxisome proliferator-activated receptor- γ in maintenance of the characteristics of mature 3T3-L1 adipocytes. *Diabetes* **51**, 2045–2055 (2002).
- Dowell, P. *et al.* Identification of nuclear receptor corepressor as a peroxisome proliferator-activated receptor alpha interacting protein. *J. Biol. Chem.* **274**, 15901–15907 (1999).
- McBurney, M. W. *et al.* The mammalian SIR2 α protein has a role in embryogenesis and gametogenesis. *Mol. Cell. Biol.* **23**, 38–54 (2003).
- Lin, S.-J., Ford, E., Haigis, M., Liszt, G. & Guarente, L. Calorie restriction extends yeast life span by lowering the level of NADH. *Genes Dev.* **18**, 12–16 (2004).
- Gabriely, I. & Barzilai, N. The role of fat cell derived peptides in age-related metabolic alterations. *Mech. Ageing Dev.* **122**, 1565–1576 (2001).
- Miles, P. D., Barak, Y., Evans, R. M. & Olefsky, J. M. Effect of heterozygous PPAR γ deficiency and TZD treatment on insulin resistance associated with age and high-fat feeding. *Am. J. Physiol.* **284**, E618–E626 (2003).
- Marette, A., Deshaies, Y., Collet, A. J., Tulp, O. & Bukowiecki, L. J. Major thermogenic defect associated with insulin resistance in brown adipose tissue of obese diabetic SHR/N-cp rats. *Am. J. Physiol.* **261**, E204–E213 (1991).
- Forman, B. M. *et al.* 15-Deoxy- Δ 12,14 prostaglandin J2 is a ligand for the adipocyte determination factor PPAR γ . *Cell* **83**, 803–812 (1995).
- Le Lay, S., Lefrère, I., Trautwein, C., Dugail, I. & Krief, S. Insulin and sterol-regulatory element-binding protein-1c (SREBP-1c) regulation of gene expression in 3T3-L1 adipocytes. *J. Biol. Chem.* **277**, 35625–35634 (2002).
- Fajas, L. *et al.* E2Fs regulate adipocyte differentiation. *Dev. Cell* **3**, 39–49 (2002).
- Langley, E. *et al.* Human SIR2 deacetylates p53 and antagonizes PML/p53-induced cellular senescence. *EMBO J.* **21**, 2383–2396 (2002).
- Dowell, P., Peterson, V. J., Zabriskie, M. & Leid, M. Ligand-induced peroxisome proliferator-activated receptor alpha conformational change. *J. Biol. Chem.* **272**, 2013–2020 (1997).

Acknowledgements We thank B. Spiegelman for the gift of constructs and for discussion; T. Kouzarides and T. Heinzel for Sirt1 and NCoR constructs, respectively; and R. Frye for the anti-Sirt1 antiserum. The Guarente laboratory was supported by grants from the NIH, the Ellison Medical Foundation, and the Howard and Linda Stern Fund. Studies in the Leid laboratory were supported by grants from the National Institutes of Health and by a NIEHS Center grant to the Oregon State University Environmental Health Sciences Center. T.S. was supported by a graduate fellowship from the Royal Thai Government. R.M.O. is supported by the Portuguese Foundation for Science and Technology. M.K. is a Howard Hughes undergraduate research fellow. F.P. was supported in part by a postdoctoral fellowship from the Canadian Institutes of Health Research.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to L.G. (leng@mit.edu).

Plant retinoblastoma homologues control nuclear proliferation in the female gametophyte

Chantal Ebel, Luisa Mariconti & Wilhelm Grissem

Institute of Plant Sciences, Swiss Federal Institute of Technology, ETH Zürich, Universitätstrasse 2, CH-8092 Zürich, Switzerland

Haploid spores of plants divide mitotically to form multicellular gametophytes. The female spore (megaspore) of most flowering plants develops by means of a well-defined programme into the mature megagametophyte consisting of the egg apparatus and a central cell^{1,2}. We investigated the role of the *Arabidopsis* retinoblastoma^{3,4} protein homologue and its function as a negative regulator of cell proliferation during megagametophyte development. Here we show that three mutant alleles of the gene for the *Arabidopsis* retinoblastoma-related protein, *RBR1* (ref. 4), are gametophytic lethal. In heterozygous plants 50% of the ovules are aborted when the mutant allele is maternally inherited. The mature unfertilized mutant megagametophyte fails to arrest mitosis and undergoes excessive nuclear proliferation in the embryo sac. Supernumerary nuclei are present at the micropylar end of the megagametophyte, which develops into the egg apparatus and central cell. The central cell nucleus, which gives rise to the endosperm after fertilization, initiates autonomous endosperm development reminiscent of *fertilization-independent seed (fis)* mutants⁵. Thus, *RBR1* has a novel and previously unrecognized function in cell cycle control during gametogenesis and in the repression of autonomous endosperm development.

During female gametophyte development in *Arabidopsis*, the haploid nucleus of one of the four meiotic spores undergoes three consecutive mitotic divisions. This is followed by cellularization, in which the eight nuclei are assembled into a seven-cell structure comprised of three antipodals, two synergids and the egg cell (egg apparatus), and a binucleate central cell. After degeneration of the three antipodal cells and fusion of the central cell nuclei, the mature female gametophyte (megagametophyte) has four visible nuclei before fertilization^{1,2}. Although mutants have been identified that affect megagametophyte development⁶, maize *indeterminate gametophyte1 (ig)* is the only known mutation that fails to arrest mitotic divisions in the mature megagametophyte⁷. In contrast, mitotic division of the central cell nucleus and autonomous endosperm development in the absence of fertilization is common to the *fis*-class of mutants^{8–12} that have a gametophytic maternal effect. *FIS*-class genes encode proteins of the polycomb group (PcG), including MEA, FIE, FIS2 and the WD-40 repeat protein MSI1, which control cell proliferation^{5,8–12}. In animals and plants, the

retinoblastoma protein (pRB) and its orthologues interact with specific PcG subunits and RbAp48 (also known as MSI1)^{13–15}. pRB is a key negative regulator of cell division in metazoans and controls the G1/S transition by repressing E2F transcription factors^{16,17}. The E2F–pRB pathway is conserved in plants^{3,18,19}, and *Arabidopsis* is currently the only higher plant known to contain a single retinoblastoma-related gene (*RBR1*)⁴. To understand the role of *RBR* in female gametophyte development, we examined *rbr* mutant alleles that showed gametophytic lethality.

A comprehensive search of T-DNA insertion mutant collections identified three *Arabidopsis* *RBR1* alleles (*rbr1-1*, *rbr1-2* and *rbr1-3*) in which the T-DNA disrupts the central region of the gene (Fig. 1a). Premature termination in either exon 3, 19 or 20 of the mouse *Rb* gene results in truncated non-functional proteins^{20–22}. It was therefore likely that the T-DNA insertions in *RBR1* were null alleles. None of the T3 plants that showed normal vegetative development was homozygous for the mutant allele. For all three mutant alleles, mature siliques contained normal, green, fertilized seeds and aborted ovules at a 1:1 ratio, suggesting gametophytic lethality (Fig. 1b, lower panel)^{6,8}. The 50% ovule abortion ratio (Table 1) always co-segregated with the T-DNA insertion. As the phenotype was identical for all three mutant alleles, the following results are representative for *rbr1-1*. Reciprocal crosses with the corresponding wild-type Col-0 (Tables 1 and 2) confirmed that *rbr1-1* could not be transmitted when wild-type pollen was used to fertilize *rbr1-1/+* (heterozygous) pistils, although paternal transmission was possible

with a low efficiency. This reduced transmission is consistent with a gametophytic defect and suggests that seed development strictly requires the maternal *Arabidopsis* *RBR1* allele but that *rbr1-1* also strongly affects male gametophyte development or function.

We examined whole-mounted, cleared ovules from *rbr1-1/+* and wild-type plants at different stages of megagametogenesis. During early stages (female gametophyte (FG) stages FG1–FG5; ref. 2) development of *rbr1-1* megagametophytes was indistinguishable from wild type. Formation of ovule inner and outer integuments as well as meiosis ($n = 199$, data not shown) proceeded normally in *rbr1-1*, indicating that the maternal sporophytic tissue was not responsible for the ovule abortion. At FG7 (ovule maturity), the three antipodal cells of the *rbr1-1* megagametophyte were visible at the chalaza, indicating that the embryo sac was correctly polarized. However, all *rbr1-1* megagametophytes (54% of gametophytes in *rbr1-1/+* carpels; $n = 224$) had supernumerary nuclei (7–15) clustered at the micropyle (Fig. 1c). The nuclei were irregular in size and appeared partially enclosed by membranes or cell-wall-like structures, suggesting that nuclei in the micropylar domain where the egg apparatus will be formed were not arrested (Fig. 1c, arrowheads). At present it is not possible to assign a clear identity to the supernumerary nuclei and cells because no marker lines are available that specify cells in the micropylar domain. It is also unknown at which stage of the cell cycle the nuclei in this domain are arrested before fertilization. However, the lack of mitotic cyclin B1 expression²³ (data not shown) in the egg apparatus would

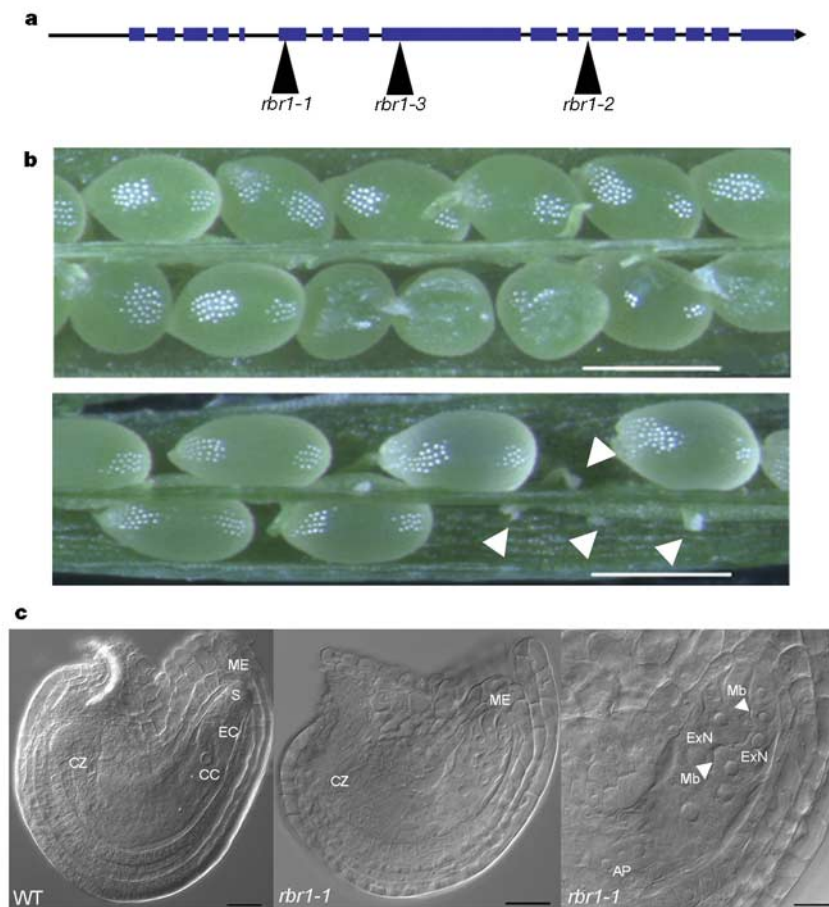


Figure 1 *Arabidopsis* *rbr1* mutant alleles induce ovule abortion and proliferation of nuclei in the egg apparatus. **a**, *RBR1* gene structure and location of T-DNA insertions in the *rbr1-1*, *rbr1-2* and *rbr1-3* mutant alleles. **b**, Open siliques from wild-type (upper panel) and *rbr1-1/+* (lower panel) plants. Arrowheads indicate aborted ovules. Scale bar, 500 μ m. **c**, Mature wild-type (WT) ovules before fertilization (FG7; left panel). A mutant female gametophyte from the same silique has supernumerary nuclei at the micropyle

(middle panel). A close-up view of the micropylar end of the *rbr1-1* megagametophyte (right panel) shows that the nuclei may be partially cellularized. Scale bars, 20 μ m (left and middle panels) and 10 μ m (right panel). AP, antipodal cells; CC, central cell; CZ, chalaza; EC, egg cell; ExN, extra nuclei; Mb, membrane or cell-wall-like structure; ME, micropylar end; S, synergid cells.

suggest that these nuclei are arrested in G1. Alternatively, the nuclei might be arrested in G2 before expression of cyclin B1. In any case, the nuclear proliferation in *rbr1-1* megagametophytes suggests that the micropylar nuclei had lost G1 control.

The interaction of *RBR1* with the WD-40 protein MSI1 (ref. 15), which is also found associated with the PcG protein complex¹², suggested that *rbr1-1* might also affect mitotic arrest of the central cell nucleus. We therefore emasculated wild-type and *rbr1-1/+* flowers and after two days prepared cleared, whole-mounted ovules from pollinated and unpollinated carpels. In unfertilized carpels, wild-type megagametophytes remained arrested at FG7 (50.2%; Fig. 2b). In addition, 27.3% ($n = 231$) of the *rbr1-1* megagametophytes in *rbr1-1/+* emasculated carpels developed a seed-like morphology with a syncytial endosperm-like domain at the chalazal pole characteristic of *fis* mutants⁵ (Fig. 2a), and an aberrantly cellularized domain at the micropyle specific to the *rbr* alleles (compare Figs 2a and 1c). The remaining 22.5% of megagametophytes had no *fis* phenotype but over-proliferating micropylar nuclei (Fig. 2c). Unlike *fis* mutants, however, *rbr1-1* megagametophytes did not proceed with seed development and aborted 4 days after emasculature. The early *fis* phenotype of *rbr1-1* megagametophytes suggests that the FIS PcG-complex-mediated arrest of the central cell nucleus before fertilization depends on *RBR1*.

We next examined whether the fertilized *rbr1-1* megagametophyte could develop into an embryo. Self-fertilized *rbr1-1* siliques contained aborted ovules with seed-like morphology and embryo-like structures (15.6% of seeds in *rbr1-1/+* siliques; $n = 501$) that had not developed beyond the pre-globular stage (Fig. 2e). Defects in embryo development were variable and included abnormal or absent suspensors, or abnormal division planes in the embryo proper. The frequency of defective embryos was similar in *rbr1-1/+* + pistils pollinated with wild-type pollen (11.7%, $n = 94$). These abnormal embryos probably result from fertilization of the *rbr1-1* megagametophyte with wild-type pollen because the *rbr1-1* paternal transmission was very low (Tables 1 and 2). Paternal *RBR1*, however, cannot rescue the maternal defect. Thus, all three *rbr* alleles show a combination of female gametophyte lethality and gametophytic maternal-effect embryo lethality, in contrast to the zygotic lethality observed for functional loss of the mouse *Rb* and *Drosophila Rbf* genes^{20–22,24}.

Because the transmission rates suggested that *Arabidopsis RBR1* function is required for both female and male gametophytes, we examined pollen viability and *RBR1* expression in the ovule. Anthers

from *rbr1-1/+* flowers had malformed dead pollen in addition to viable pollen (Fig. 2h). Affymetrix GeneChip analysis of microgametogenesis (NASC microarray database)²⁵ revealed elevated *RBR1* expression in uninucleate microspores but not in tricellular pollen. Thus, *rbr1-1* markedly decreases gametophyte viability, but unlike the effect of *rbr1-1* on megagametophyte arrest, reduction of pollen viability is not fully penetrant during microgametogenesis. Analysis of *RBR1* expression in wild-type flowers at 0, 1–3 and 4–5 days after pollination showed high transcript levels before and after fertilization (Fig. 3a). Ovule *in situ* hybridization confirmed the expression of *RBR1* in both sporophytic tissue and the megagametophyte (Fig. 3b).

Our results indicate that the arrest of the mature unfertilized megagametophyte and development of the microgametophyte in

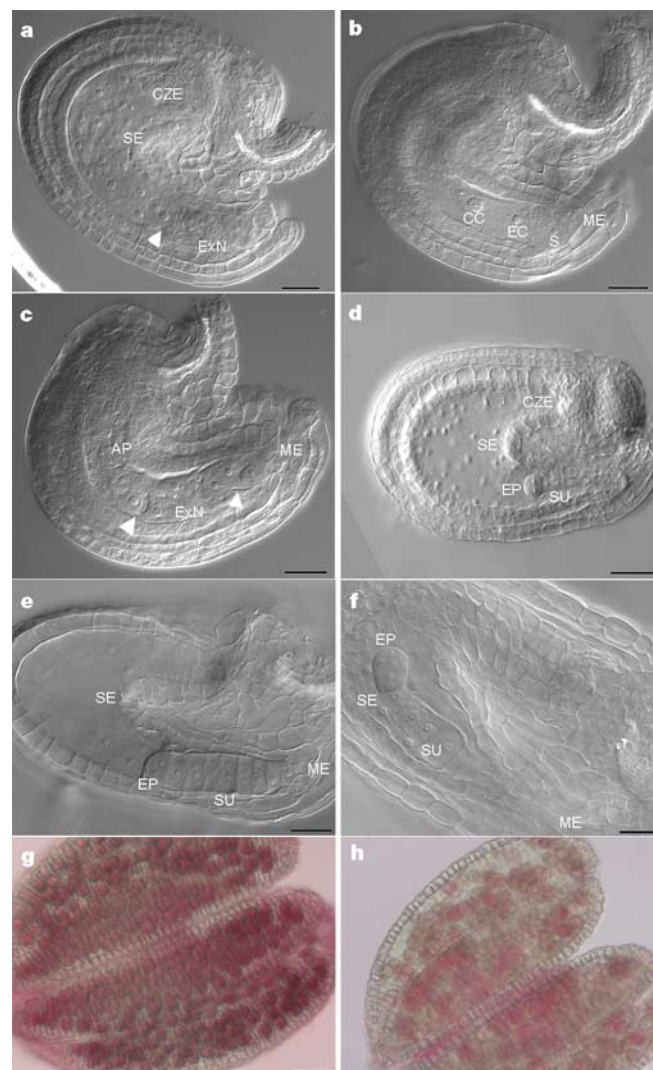


Figure 2 The *rbr1-1* mutant initiates a fertilization-independent endosperm and affects embryo and pollen development. **a**, A *rbr1-1* ovule two days after emasculature in which the central cell has initiated endosperm development in the absence of fertilization. Note that the embryo sac is clearly separated (arrowhead) into the syncytial endosperm, and the supernumerary nuclei at the micropyle (compare with Fig. 1c). **b**, Wild-type unfertilized ovule 2 days after emasculature. **c**, A second class of *rbr1-1* mutants show excessive proliferation of nuclei at the micropyle 2 days after emasculature but no autonomous endosperm development. **d**, Wild-type fertilized seed at 2 DAP with the developing pre-globular embryo. **e**, **f**, Examples of *rbr1-1* embryos with abnormal development in an aborted seed-like structure. Scale bars, 20 μ m. **g**, **h**, Alexander staining of *rbr1-1/+* (**g**) and wild-type (**h**) anthers. Viable pollen grains are stained purple whereas dead pollen grains appear greenish. Scale bars, 50 μ m. Abbreviations are as in Fig. 1 except for: EP, embryo proper; SE, syncytial endosperm; SU, suspensor.

Table 1 Percentage of aborted ovules in *rbr* mutants

Parental genotypes (female X male)	Normal	Aborted	Expected*	P-value†
<i>RBR1/RBR1</i> X <i>RBR1/RBR1</i>	1,291	29 (2%)	NA	NA
<i>rbr1-1/RBR1</i> X <i>rbr1-1/RBR1</i>	357	326 (47.7%)	341.5 (50%)	0.2367
<i>rbr1-2/RBR1</i> X <i>rbr1-2/RBR1</i>	270	271 (50.1%)	270.5 (50%)	0.9661
<i>rbr1-3/RBR1</i> X <i>rbr1-3/RBR1</i>	160	151 (48.5%)	155.5 (50%)	0.6101
<i>rbr1-1/RBR1</i> X <i>RBR1/RBR1</i>	116	114 (49.6%)	115 (50%)	0.8944
<i>RBR1/RBR1</i> X <i>rbr1-1/RBR1</i>	162	21 (11.5%)	NA	NA

Mature siliques of wild-type and *rbr1-1/+* plants were opened and scored for the ovule abortion phenotype. NA, not applicable.

*Fraction of aborted ovules in case of female gametophyte lethal mutation.

†Probability of χ^2 for a normal distribution with a degree of freedom (d.f.) = 1.

Table 2 Transmission efficiency of the *rbr1-1* allele

Parental genotypes (female X male)	Genotype of F ₁ plants		
	<i>rbr1-1/+</i>	<i>RBR1/RBR1</i>	Transmission efficiency*
<i>rbr1-1/RBR1</i> X <i>rbr1-1/RBR1</i>	35	355	NA
<i>rbr1-1/RBR1</i> X <i>RBR1/RBR1</i>	0	134	0%
<i>RBR1/RBR1</i> X <i>rbr1-1/RBR1</i>	12	157	7.6%

Transmission was scored as the number of plants having the mutant allele versus the wild-type allele determined by PCR using allele-specific primers for *rbr1-1* as described in Methods. NA, not applicable.

* (Number of mutant plants/number of wild-type plants)X100%, through female or male gametes as defined in ref. 6.

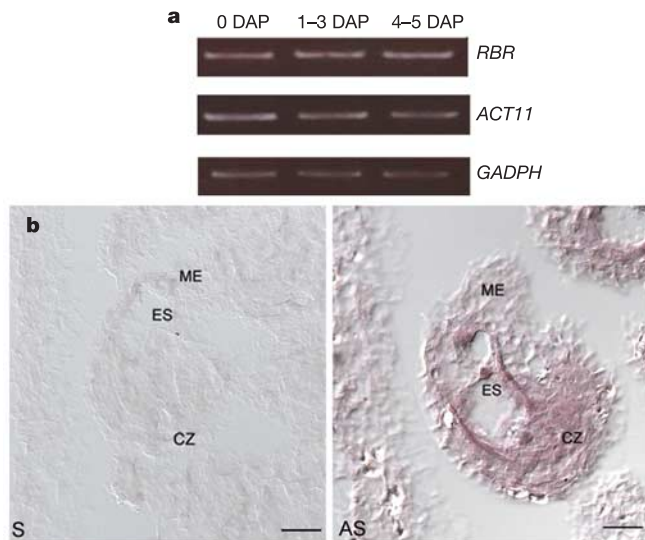


Figure 3 *RBR1* is expressed in the female gametophyte and at different stages of flower development. **a**, PCR with reverse transcription (RT-PCR) analyses of *RBR1* expression (upper panel; 35 cycles) in a wild-type flower before fertilization (0 DAP), in silicles containing embryos at pre-globular stages (1–3 DAP) and at late globular stages (4–5 DAP). *GAPDH* and *ACTIN11* were used as control (middle and lower panels; 30 cycles). **b**, Localization of *RBR1* messenger RNA in wild-type megagametophytes before fertilization. Sections were hybridized with sense (left) and antisense (right) *RBR1* probes. Abbreviations are as in Fig. 1 except for: AS, antisense; S, sense.

Arabidopsis depend on *RBR1*. Plant germline cells originate late in development from somatic diploid tissue that previously generated all other post-embryonic aerial organs of the plant. The alternation of generations and formation of multicellular haploid gametophytes are peculiar to plant development, and allowed us to discover a novel function of the pRB plant homologue in the female gametophyte. Unlike plants, metazoans develop reproductive organs during embryogenesis in which stem cells are set aside that produce germline cells for the remaining lifespan of the organism. Although pRB and its homologues p107 and p130 are differentially expressed and phosphorylated during spermatogenesis²⁶, no defects in gametogenesis have been reported for the loss of function of the pRB family members^{20–22}. Similarly, *Drosophila* *Rbf* loss-of-function mutants can form viable female gametes, but embryos die with ectopic S-phase cells and increased apoptosis²⁴. Thus, *Arabidopsis* *RBR1* has an earlier function involved in female gametophyte development before fertilization and embryogenesis.

Although the precise molecular mechanism that controls megagametophyte arrest remains to be established, our results suggest that *Arabidopsis* *RBR1* is required upstream of FIS or together with the MEA–FIE PcG complex. The corresponding human PRC2 complex acts downstream of the pRB–E2F pathway²⁷. In plants MSI1 interacts with the MEA–FIE PcG complex and also with *RBR1* (refs 12, 15). Thus, MSI1 could link *RBR1* and the PcG complex to arrest proliferation of the central cell by repressing genes essential for endosperm development in an E2F-independent manner. However, *Arabidopsis* *fis* mutants do not proliferate supernumerary nuclei at the micropyle^{8–12}, suggesting that arrest of the egg apparatus is controlled by *RBR1* through a FIS-independent mechanism. Furthermore, whereas *fis* mutants develop over-proliferating embryos between the globular and heart stages^{8,12}, fertilized *rbr1-1* megagametophytes rarely form embryos and these do not proceed to the globular stage. Thus, the *Arabidopsis* *rbr* alleles have uncovered novel *RBR* functions in the mitotic proliferation of nuclei in the micropylar domain, which includes the egg apparatus, and the epigenetically controlled arrest of the central cell nucleus before double fertilization. □

Methods

Plant lines and growth conditions

Arabidopsis plants used in our study were all derived from the Col-0 accession line. The *rbr1-1* (SALK_012270) and *rbr1-2* alleles (SALK_002946) were obtained from the SALK T-DNA insertion collection (<http://signal.salk.edu>). The *rbr1-3* (GABI_170G02) allele was obtained from the GABI-Kat T-DNA mutant collection (http://www.mpiz-koeln.mpg.de/GABI-Kat/GABI-Kat_homepage.html; ref. 28). Seeds were surface-sterilized using 5% bleach and germinated on MS medium. After two weeks, the seedlings were transferred to soil and grown in Conviron chambers with a 16/8 h light/dark cycle at 23 °C in 70% humidity.

Genomic DNA isolation and PCR screening

To identify plants with the T-DNA insertion in *RBR1* we performed polymerase chain reaction (PCR) analyses because the kanamycin-resistance gene (*NPTR*) was silenced. Allele-specific PCR reactions were performed to confirm the T-DNA insertion sites using primers *LBb1* (5′-GCGTGGACCGCTTGTGCAACT-3′) or *LB-GABI* (5′-ATATTGACCATCATCTCATTGC-3′) and the following specific primers for the mutant alleles: *RBR F4* (5′-CAGCTGCGTTGGAGACTATGA-3′) for *rbr1-1*, *RBR R10* (5′-GCTGGAGTAGTAAGTTATTCT-3′) for *rbr1-2*, and *RBR R4* (5′-CAGTACCAATTCAGCTGAGCA-3′) for *rbr1-3*.

Whole-mount preparation

Flowers and pistils of different developmental stages were prepared as described¹². Briefly, the tissue was fixed in ethanol-acetic acid (9:1) overnight at 4 °C and dehydrated by two subsequent 1 h steps in 90% and 70% ethanol. The preparation was then cleared in chloral hydrate (chloral hydrate:water:glycerol (8:2:1)) overnight at 4 °C. Pistils were dissected and observed using DIC optics (Zeiss). Images were recorded with an Axiocam HRC CCD camera (Zeiss).

In situ hybridization

Expression of *Arabidopsis* *RBR1* was analysed by *in situ* hybridization as described earlier²⁹ with some modifications. For sense and antisense 11-digoxigenin-UTP probe synthesis, a full-length *Arabidopsis* *RBR1* complementary DNA clone in pSK was linearized using *XhoI* and *SacII* and used for probe synthesis. The labelled probes were then hydrolysed in carbonate buffer (120 mM Na₂CO₃, 80 mM NaHCO₃) for 27 min at 60 °C, followed by ethanol precipitation. Mature wild-type flowers at 0 days after pollination (DAP) were fixed in 4% paraformaldehyde and embedded in Paraplast X-plus (Sigma). Twelve-micrometre sections were prepared with a Leica RM2145 microtome and mounted on polysine slides (Menzel-Gläser). Sections were then digested with proteinase K for 30 min at 37 °C, treated with acetic acid anhydride, dried in ethanol and hybridized with 50 ng of DIG-labelled probe per slide overnight at 55 °C, followed by washing with × 0.2 SSC at 55 °C. The slides were then blocked with DIG blocking reagent, followed by incubation with an anti-DIG antibody that was alkaline-phosphatase-conjugated (Roche). After washing with blocking reagent the detection was performed in the presence of NBT, X-phosphate and levamisole (1 mM) for 16–48 h. Reactions were stopped with TE buffer (10 mM, pH 8.0) and mounted in TE/glycerol (v/v) for microscopic analysis.

Alexander staining

Flowers were collected in 10% ethanol and incubated overnight at 4 °C. Dissected anthers were incubated with Alexander stain³⁰ under the coverslip for 15 min at room temperature. Observation was performed with a Zeiss microscope using DIC optics (Zeiss).

RNA isolation and RT-PCR

Total RNA was isolated from wild-type flowers at 0, 1–3 and 4–6 DAP using the TRIzol reagent (Invitrogen). Five-hundred nanograms of total RNA was digested with RNase-free DNase (Promega Biosciences) and used for reverse transcription with Superscript II (Invitrogen). After 1/10 dilution, 1 μl of the synthesized cDNA was used for PCR using *Arabidopsis* *RBR1*-specific primers spanning an intron to control for DNA contamination: *RBR F1* (5′-ATCTCCAGCATCTGAGTGCC-3′) and *RBR R1* (5′-TATGACAGTCTGAGCCACT-3′). *GAPDH-F* (5′-TTCTTGGCACCAGCTTCAAT-3′) and *GAPDH-R* (5′-CTCCCTTGGAAAGGAGCTAGG-3′) as well as *ACT11-F* (5′-GGAACAGTGTGACTCACACCATC-3′) and *ACT11-R* (5′-AAGCTGTTCTTCCCTCTACG-3′) were used as controls following conditions as described in ref. 12.

Received 4 March; accepted 10 May 2004; doi:10.1038/nature02637.

- Grossniklaus, U. & Schneitz, K. The molecular and genetic basis of ovule and megagametophyte development. *Semin. Cell Dev. Biol.* **9**, 227–238 (1998).
- Drews, G. N. & Yadegari, R. Development and function of the angiosperm female gametophyte. *Annu. Rev. Genet.* **36**, 99–124 (2002).
- Durfee, T., Feller, H. S. & Gruissem, W. Retinoblastoma-related proteins in plants: homologues or orthologues of their metazoan counterparts? *Plant Mol. Biol.* **43**, 635–642 (2000).
- Kong, L. J. *et al.* A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *EMBO J.* **19**, 3485–3495 (2000).
- Grossniklaus, U., Spillner, C., Page, D. R. & Kohler, C. Genomic imprinting and seed development: endosperm formation with and without sex. *Curr. Opin. Plant Biol.* **4**, 21–27 (2001).
- Howden, R. *et al.* Selection of T-DNA-tagged male and female gametophytic mutants by segregation distortion in *Arabidopsis*. *Genetics* **149**, 621–631 (1998).
- Lin, B. Y. Structural modifications of the female gametophyte associated with *indeterminate gametophyte* (*ig1*) mutant in maize. *Can. J. Genet. Cytol.* **20**, 249–257 (1978).
- Grossniklaus, U., Vielle-Calzada, J. P., Hoepfner, M. A. & Gagliano, W. B. Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* **280**, 446–450 (1998).
- Ohad, N. *et al.* A mutation that allows endosperm development without fertilization. *Proc. Natl Acad.*

- Sci. USA **93**, 5319–5324 (1996).
10. Kiyosue, T. *et al.* Control of fertilization-independent endosperm development by the MEDEA polycomb gene in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **96**, 4186–4191 (1999).
 11. Luo, M. *et al.* Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **96**, 296–301 (1999).
 12. Köhler, C. *et al.* *Arabidopsis* MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO J.* **22**, 4804–4814 (2003).
 13. Lu, X. & Horvitz, H. R. lin-35 and lin-53, two genes that antagonize a *C. elegans* Ras pathway, encode proteins similar to Rb and its binding protein RbAp48. *Cell* **95**, 981–991 (1998).
 14. Dahiya, A., Wong, S., Gonzalo, S., Gavin, M. & Dean, D. C. Linking the Rb and polycomb pathways. *Mol. Cell* **8**, 557–569 (2001).
 15. Ach, R. A., Taranto, P. & Gruijssem, W. A conserved family of WD-40 proteins binds to the retinoblastoma protein in both plants and animals. *Plant Cell* **9**, 1595–1606 (1997).
 16. Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
 17. Harbour, J. W. & Dean, D. C. The Rb/E2F pathway: expanding roles and emerging paradigms. *Genes Dev.* **14**, 2393–2409 (2000).
 18. Ach, R. A. *et al.* RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein. *Mol. Cell. Biol.* **17**, 5077–5086 (1997).
 19. De Veylder, L. *et al.* Control of proliferation, endoreduplication and differentiation by the *Arabidopsis* E2Fa-DPa transcription factor. *EMBO J.* **21**, 1360–1368 (2002).
 20. Lee, E. Y. *et al.* Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis. *Nature* **359**, 288–294 (1992).
 21. Jacks, T. *et al.* Effects of an Rb mutation in the mouse. *Nature* **359**, 295–300 (1992).
 22. Clarke, A. R. *et al.* Requirement for a functional Rb-1 gene in murine development. *Nature* **359**, 328–330 (1992).
 23. Kwee, H. S. & Sundaresan, V. The NOMEA gene required for female gametophyte development encodes the putative APC6/CDC16 component of the Anaphase Promoting Complex in *Arabidopsis*. *Plant J.* **36**, 853–866 (2003).
 24. Du, W. & Dyson, N. The role of RBF in the introduction of G1 regulation during *Drosophila* embryogenesis. *EMBO J.* **18**, 916–925 (1999).
 25. Honys, D. & Twell, D. Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* **132**, 640–652 (2003).
 26. Yan, W., Kero, J., Suominen, J. & Toppari, J. Differential expression and regulation of the retinoblastoma family of proteins during testicular development and spermatogenesis: roles in the control of germ cell proliferation, differentiation and apoptosis. *Oncogene* **20**, 1343–1356 (2001).
 27. Bracken, A. P. *et al.* EZH2 is downstream of the PRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* **22**, 5323–5335 (2003).
 28. Li, Y., Rosso, M. G., Strizhov, N., Viechover, P. & Weisshaar, B. GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in *Arabidopsis thaliana*. *Bioinformatics* **19**, 1441–1442 (2003).
 29. Vielle-Calzada, J. P. *et al.* Maintenance of genomic imprinting at the *Arabidopsis* medea locus requires zygotic DDM1 activity. *Genes Dev.* **13**, 2971–2982 (1999).
 30. Alexander, M. P. Differential staining of aborted and nonaborted pollen. *Stain Technol.* **44**, 117–122 (1969).

Acknowledgements We are grateful to M. Collinge, J. Fütterer, U. Grossniklaus, M. Hanin and L. Hennig for discussions and critical reading of the manuscript. We thank J. Wyrzykowska and A. Johnston for help with *in situ* hybridization and ovule clearings. We also thank the Salk Institute Genomic Analysis Laboratory for providing the sequence-indexed *Arabidopsis* T-DNA insertion mutants and ABRC for providing us with seeds. This work was supported by funding from the Swiss National Science Foundation to W.G.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.G. (wilhelm.gruijssem@ipw.biol.ethz.ch).

Structural basis of protein phosphatase 1 regulation

Mohammed Terrak*, Frederic Kerff*, Knut Langsetmo, Terence Tao & Roberto Dominguez

Boston Biomedical Research Institute, 64 Grove Street, Watertown, Massachusetts 02472, USA

* These authors contributed equally to this work

The coordinated and reciprocal action of serine/threonine (Ser/Thr) protein kinases and phosphatases produces transient phosphorylation, a fundamental regulatory mechanism for many biological processes. The human genome encodes a far greater number of Ser/Thr protein kinases than of phosphatases. Protein phosphatase 1 (PP1), in particular, is ubiquitously distributed and regulates a broad range of cellular functions, including

glycogen metabolism, cell-cycle progression and muscle relaxation^{1,2}. PP1 has evolved effective catalytic machinery but lacks substrate specificity. Substrate specificity is conferred upon PP1 through interactions with a large number of regulatory subunits. The regulatory subunits are generally unrelated, but most possess the RVxF motif, a canonical PP1-binding sequence. Here we reveal the crystal structure at 2.7 Å resolution of the complex between PP1 and a 34-kDa N-terminal domain of the myosin phosphatase targeting subunit MYPT1. MYPT1 is the protein that regulates PP1 function in smooth muscle relaxation³. Structural elements amino- and carboxy-terminal to the RVxF motif of MYPT1 are positioned in a way that leads to a pronounced reshaping of the catalytic cleft of PP1, contributing to the

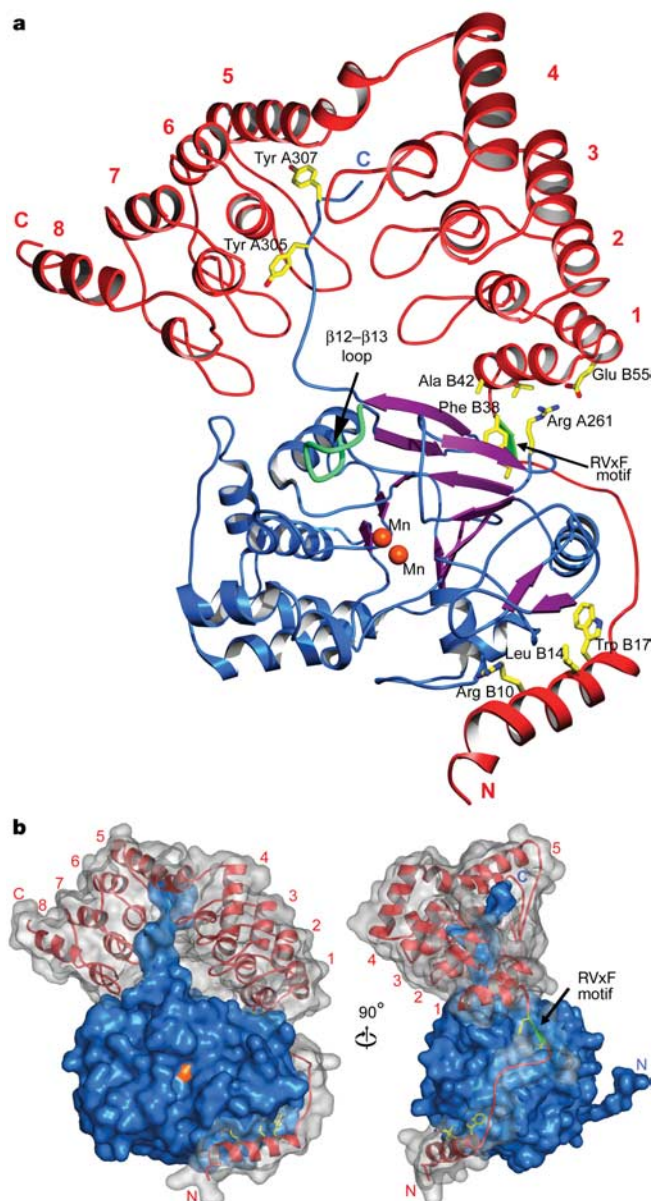


Figure 1 General fold of the PP1δ-MYPT1₁₋₂₉₉ complex. **a**, Ribbon representation: PP1δ; α-helices and loops (blue) and β-strands (magenta); and MYPT1 (red). The two cations in the catalytic site are coloured orange. The ankyrin repeats of MYPT1 are numbered 1 to 8 from the N to the C terminus. The major contacts with PP1 involve three separate regions of MYPT1: the N-terminal arm, the RVxF motif and the second group of ankyrin repeats, which interact mainly with PP1 residues Tyr A305 and Tyr A307. **b**, Two surface representations of the complex, rotated by 90 degrees (PP1δ, blue; MYPT1, red).

increased myosin specificity of this complex. The structure has general implications for the control of PP1 activity by other regulatory subunits.

In smooth muscle and non-muscle cells, phosphorylation of Ser 19 of the myosin regulatory light chain (RLC) turns on the ATPase activity of myosin. Myosin light chain kinase (MLCK) and myosin phosphatase catalyse the reactions of phosphorylation and dephosphorylation, regulating muscle contraction and relaxation, respectively. Myosin phosphatase is a trimeric holoenzyme, comprising a catalytic subunit (PP1, δ -isoform), a 130-kDa targeting subunit (MYPT1) and a 20-kDa subunit (M20) of unknown function³. MYPT1 plays a key part in determining the physical and functional integrity of the trimeric myosin phosphatase holoenzyme. It binds PP1 δ at the N terminus^{4–8} and M20 at the C terminus^{3,4}. Binding to myosin seems to take place through two separate sites: a site situated within the C-terminal one-third of MYPT1 (ref. 9), and a site responsible for catalysis that engages PP1 and the N-terminal one-third of MYPT1 simultaneously. In a similar way to that of the intact myosin phosphatase holoenzyme, the complex between PP1 and the N-terminal one-third of MYPT1 displays ~ 15 -fold increased catalytic activity and ~ 10 -fold higher affinity for phosphorylated myosin (or P-RLC) than the isolated catalytic subunit^{4–7,10}. In addition, this complex has reduced activity towards other substrates, such as phosphorylase *a* and glycogen synthase^{4,5,7}. Together these findings suggest that different parts of MYPT1 are responsible for substrate specificity and targeting.

Here we reveal the 2.7-Å-resolution structure of the complex between chicken gizzard PP1 δ and amino acids Met 1 to Ser 299 of MYPT1 (MYPT1_{1–299}) (see Methods). Crystal structures have been determined previously for complexes of PP1 (α and γ isoforms) with three natural toxin inhibitors: microcystin¹¹, okadaic acid¹² and calyculin A (ref.13), with tungstate¹⁴ and a peptide from the muscle glycogen-targeting regulatory subunit (G_M)¹⁵. Despite the differences in PP1 isoforms, crystallization conditions and crystal packing contacts, these structures are very similar to one another and to the structure obtained here for PP1 δ –MYPT1_{1–299}. The overall fold of PP1 consists of two tightly linked domains, an N-terminal α/β domain (amino acids A1–A160; chain A designates PP1 δ) and a C-terminal β domain (A161–A327), which also contains three α -helices. Most of the β -strands from the two domains converge at the inter-domain interface, where they form

a β -sandwich made of two mixed β -sheets (Fig. 1). The catalytic site lies within a large Y-shaped cleft at the inter-domain interface^{11–13}. The three branches of the Y-shaped cleft are commonly referred to as the hydrophobic, acidic and C-terminal grooves (Fig. 2a). Separating the acidic and the C-terminal grooves is a protruding loop (Asn 271 to Asp 277), which connects β -strands β 12 and β 13. This loop is ideally positioned to interact with PP1 substrates and inhibitors^{11–13} and has been linked with inhibitor specificity¹⁶. There are two metal ions in the catalytic site of the PP1 δ –MYPT1_{1–299} structure. By analogy with previous structures^{11–13}, and because of the presence of Mn²⁺ in the purification and crystallization buffers, we have identified these ions as Mn²⁺. Note that the addition of Mn²⁺ is required for full activity of this complex. However, we cannot rule out that one of the two cations might be Fe²⁺, as in the PP1 γ -tungstate structure¹⁴.

The most noticeable differences between the current and previous structures of PP1 (refs 11–15) occur at the N- and C-terminal ends, where the sequences of PP1 isoforms diverge the most. Previous structures of PP1 α (ref. 11) and PP1 γ (refs 12–15) typically begin with Leu 7 and end at Ala 299 (Ser 299 in PP1 δ); that is, 25 to 30 amino acids are disordered at the C termini of these structures. In the current structure Glu A300 to Gly A309 are trapped between the two groups of ankyrin repeats of MYPT1 and, as a result, are ordered. Thus, this structure suggests that some regulatory subunits may interact with the divergent and flexible C termini of PP1 to mediate isoform specificity.

MYPT1 interacts with PP1 over a large (2,560 Å²) intermolecular interface. This interaction, involving primarily the C-terminal β -domain of PP1, does not lead to any noticeable structural change in PP1, other than the partial stabilization of the C-terminal end. The absence of an allosteric effect may be unexpected, yet MYPT1 has a profound effect on the global shape and charge distribution of the catalytic cleft of PP1 (Fig. 2). To understand this point better we divide MYPT1 conceptually into three functional domains: the N-terminal arm (amino acids B1–B34; chain B designates MYPT1), the RVxF motif (amino acids B35–B38) and the ankyrin repeat domain (amino acids B39–B299).

The 34 N-terminal amino acids of MYPT1 preceding the RVxF motif form a long arm that wraps around PP1 to reach the base of the Y-shaped catalytic cleft (Figs 1 and 2). Part of this arm folds into an α -helix (Met B3 to Gly B19). Hydrophobic interactions involving

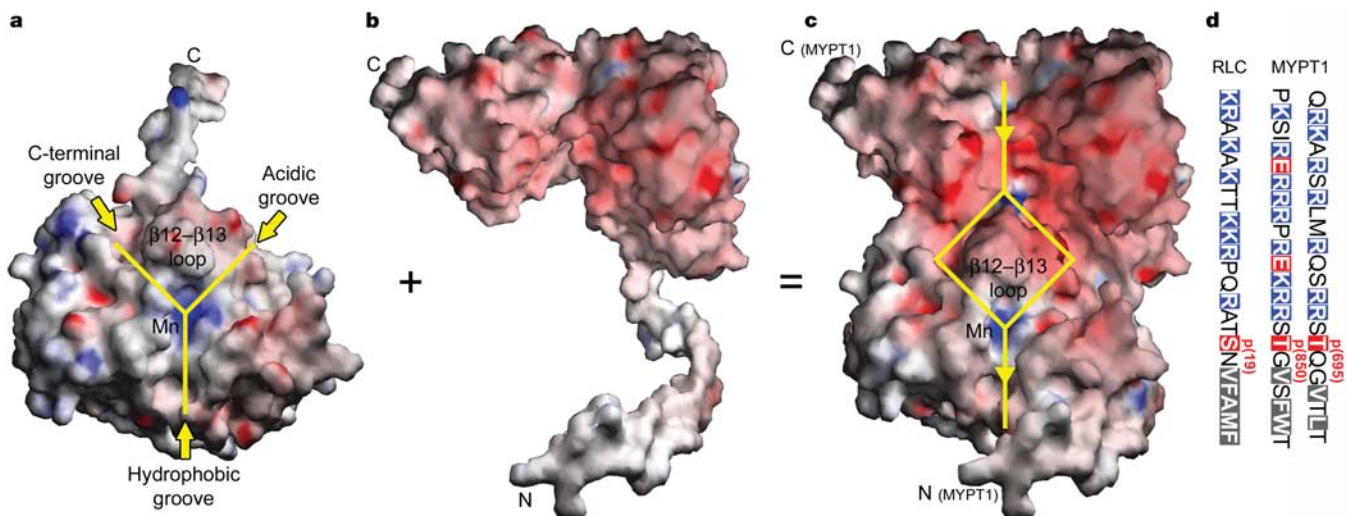


Figure 2 Electrostatic surface representations of the PP1 δ –MYPT1_{1–299} complex. **a–c**, Surfaces of PP1 (**a**), MYPT1 (**b**) and their complex (**c**) calculated using identical parameters (red and blue indicate regions charged negatively and positively, respectively). The Y-shaped catalytic cleft of PP1 is composed of three grooves: hydrophobic, acidic and C-terminal. The β 12– β 13 loop separates the acidic and the C-terminal grooves. The

reshaping of the catalytic cleft of PP1 results from the binding of the N terminus of MYPT1 near the hydrophobic groove and the addition of the acidic cleft of the ankyrin repeats at the other end of the cleft. **d**, Sequences of the RLC around Ser 19 and the MYPT1 around the regulatory phosphorylation sites Thr 695 and Thr 850. Note the existing charge complementarity between these sequences and the catalytic cleft of myosin phosphatase.

residues in this α -helix and a hydrophobic pocket in PP1 determine the position of the N-terminal arm (Fig. 3a).

The location of the N terminus of MYPT1 in the structure fits the biochemical data well. A peptide corresponding to the sequence Met B1 to Phe B38 of MYPT1 can confer upon PP1 some of the myosin specificity and increased activity characteristic of the myosin phosphatase holoenzyme^{4,6–8}. However, a shorter peptide (Asp B23 to Phe B38) binds to PP1 but has no effect on its catalytic activity⁷. Together these results emphasize the importance of the N terminus of MYPT1 for the myosin specificity of myosin phosphatase. Whereas the structure negates the existence of an allosteric effect on PP1, it places the N terminus of MYPT1 near the hydrophobic groove of the catalytic cleft, where it may interact with the myosin substrate directly (steric effect). Hydrophobic residues at the N terminus of MYPT1 (Met B1, Met B3 and Ala B6) and the side chain of Lys B2, add to the general hydrophobicity of this branch of the catalytic cleft of PP1 (Fig. 2). Hydrophobic amino

acids C-terminal to Ser 19 of the myosin RLC are likely to interact within this part of the cleft, making contacts with the N terminus of MYPT1 (Fig. 2d).

The numerous PP1 regulatory subunits share no significant sequence similarity, but most possess the RVxF motif, a specific PP1 targeting sequence². The crystal structure of PP1 γ complexed with a RVxF-containing peptide from the G_M subunit reveals six amino acids (⁶³RRVSEFA⁶⁹) bound in a hydrophobic pocket distal to the catalytic site¹⁵. In what seems to be a similar manner, the side chains of the Val and Phe residues in the RVxF motif of MYPT1 (³⁵KVKF³⁸) are embedded in the same pocket in PP1 δ (Fig. 3c, d). Residues Val B36 to Asp B39 adopt an extended conformation and are incorporated as an additional β -strand to a β -sheet in PP1, parallel to the N-terminal half of β 14. Also making up part of this cluster of interactions are the side chains of Ala B42 and Val B43 from the first ankyrin repeat of MYPT1. Finally, Asp A242 of PP1 makes two hydrogen-bonding contacts with the main-chain nitro-

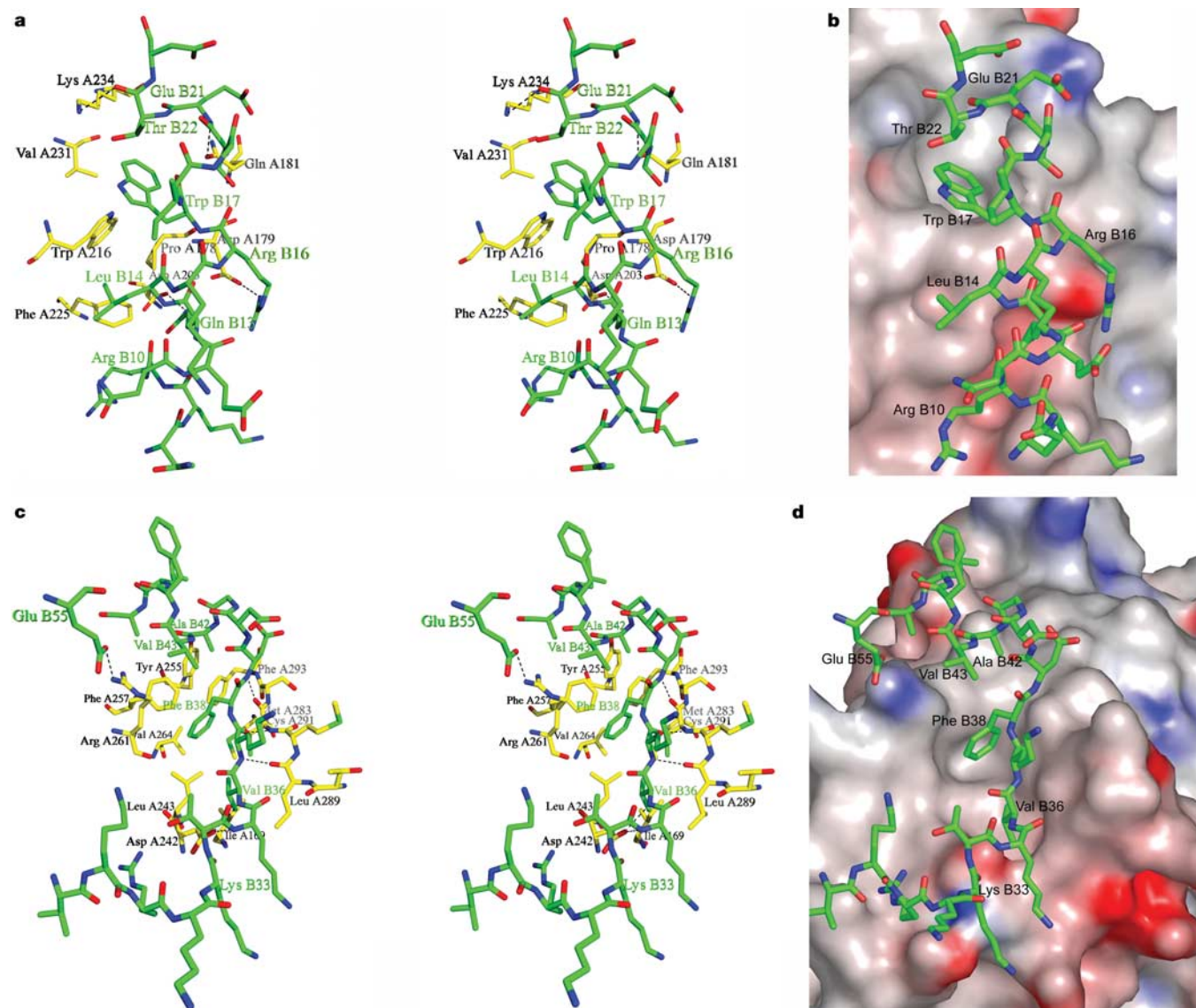


Figure 3 Interactions between PP1 and MYPT1. **a**, Stereo diagram of the interactions between the N-terminal arm of MYPT1 (green carbon atoms) and PP1 (yellow carbon atoms). **b**, The same region of MYPT1 superimposed onto an electrostatic surface representation of PP1, where red, blue and white indicate negatively-charged, positively-charged and non-charged regions, respectively. **c**, Stereo diagram of the interactions

involving the RVxF motif and the first α -helix of ankyrin repeat 1. **d**, The same region of MYPT1 superimposed onto an electrostatic surface representation of PP1. Note that the basic sequence preceding the RVxF motif does not interact directly with the acidic patch at the lower right corner of this figure.

gen atoms of Lys B35 and Val B36 of MYPT1. The ensemble of these interactions represents the single most important contribution to the formation of the PP1-MYPT1 complex. Other interactions observed in this structure are comparatively weaker and can be expected to be specific for this complex. As such, these interactions are primarily relevant to the modulation of PP1 activity.

In MYPT1, as in other PP1 regulatory subunits, positively charged amino acids at the N terminus (³⁰KRKK³³) and negatively charged amino acids at the C terminus (³⁹DD⁴⁰) flank the RVxF motif. It has been suggested¹⁵ that the basic region preceding the RVxF motif may interact with a negatively charged patch near the RVxF-binding pocket (Fig. 3d). We do not observe such an interaction. The amino acids N-terminal to the RVxF motif in MYPT1 follow a path away from the negatively charged patch, making only a few interactions with PP1, mostly through main-chain atoms.

Amino acids B39–B291 of MYPT1 fold into two groups of four ankyrin repeats, separated by a hinge at Glu B172 (Fig. 1 and Supplementary Fig. 1). Ankyrin repeats are ~33-amino-acid-long motifs consisting of pairs of antiparallel α -helices connected by β -hairpin loops¹⁷. In MYPT1, repeats 1, 2, 3, 6 and 7 are canonical with the sole exception that repeat 1 forms an extra helical turn at the N terminus. Repeats 4 and 8 lack the β -hairpin loop and repeat 5 lacks the first α -helix. The interactions with PP1 are limited to repeats 1, 5, 6 and 7 (Fig. 1). Repeat 1 interacts with the RVxF-binding pocket through hydrophobic amino acids located within the extra helical turn at the N terminus (Fig. 3c). A salt bridge between Glu B55 and PP1 residue Arg A261 also involves repeat 1. The remaining interactions are almost exclusively between the second group of ankyrin repeats and the C terminus of PP1, involving residues Tyr A305 and Tyr A307 in particular. Because sequence variations between PP1 isoforms cluster towards the C termini, a possible outcome of this interaction is to determine PP1 isoform specificity. Indeed, PP1 γ lacks both Tyr residues. In PP1 α , only Tyr A305 is present, although shifted by one amino acid.

The ankyrin repeat is a widespread motif, generally associated with a protein–protein interaction function¹⁷. The present structure offers an interesting twist: the ankyrin repeats of MYPT1 seem to play only a limited part in binding, but a potentially key part in modulating the catalytic activity of PP1. This finding is in agreement with biochemical evidence that shows almost no binding to PP1 of MYPT1 constructs lacking the N-terminal 38 amino acids^{4–8}. Constructs, however, that add the ankyrin repeat domain to the N terminus of MYPT1 display increased myosin specificity and reduced activity towards other substrates, comparable to full-length myosin phosphatase^{4–6,8}.

Substrate binding to the isolated ankyrin repeat domain is also weak⁵. This may reflect the fact that binding to myosin depends on a three-dimensional organization that can only exist when MYPT1 binds to PP1. Indeed, the β -hairpin loops from the two groups of ankyrin repeats roughly face each other in the structure, forming a clamp-like structure that closes around the C terminus of PP1. This organization gives rise to a large acidic cleft on one side of the interface between the two groups of ankyrin repeats. In the structure, this cleft is positioned as an extension to the catalytic cleft of PP1 (Fig. 2). The catalytically important β 12– β 13 loop of PP1¹⁶ forms a protrusion right in the centre of this enlarged cleft.

The major outcome from the formation of the PP1-MYPT1 complex is to produce an extended catalytic cleft specifically adapted for the myosin substrate and, at the same time, less compatible with other substrates. Analysis of the sequence of the myosin RLC around the important phosphorylation site Ser 19 makes this point fully apparent (Fig. 2d). Fifty per cent of the sequence N-terminal to Ser 19 is charged positively and can be expected to bind within the expanded acidic groove of PP1 δ -MYPT1. The sequence C-terminal to Ser 19 is mainly hydrophobic

and probably binds at the hydrophobic groove, which has also been adapted for this function by the proximity of the N terminus of MYPT1. This hypothesis assumes that the directionality of the polypeptide chain of the bound substrate is from the acidic to the hydrophobic groove. This turns out to be the directionality of the auto inhibitory peptide of the structurally related phosphatase PP2B (calcineurin)¹⁸. The sequences of two regulatory phosphorylation sites in MYPT1 (Thr 695 and Thr 850) resemble that of the RLC around Ser 19. Therefore, these sequences are also good candidates to bind at the catalytic cleft of PP1 δ -MYPT1 (Fig. 2d). Whereas phosphorylation of Thr 695 inhibits myosin phosphatase¹⁹, that of Thr 850 has so far been linked with an effect on myosin binding but not on the catalytic activity⁹.

Among the regulatory subunits of PP1, p53BP2 is the only one where ankyrin repeats follow the RVxF motif, as in MYPT1. Because p53BP2 binds with higher affinity to PP1 than to the tumour suppressor p53 (ref. 20) and inhibition of PP1 is linked to phosphorylation of p53 and apoptosis²¹, it is possible that p53BP2 provides a structural link between PP1 and p53. This interaction can now be modelled by superimposing the ankyrin repeats of p53BP2, from the structure of its complex with p53 (ref. 22), onto those of MYPT1.

Analysis of the sequences of various regulatory subunits reveals important inhibitory/modulatory sequences ~34 amino acids N-terminal or ~22 amino acids C-terminal to the RVxF motif. Because 34 amino acids N-terminal to the RVxF motif allow MYPT1 to reach the catalytic cleft of PP1, other regulatory subunits may follow a similar path, including the nuclear inhibitor of PP1 (NIPP1)²³, spinophilin and neurabin²⁴ (Fig. 4). Amino acids 143–217 of NIPP1 constitute a potent PP1 inhibitor²³. This region contains regulatory phosphorylation sites at Ser 178, Ser 199,

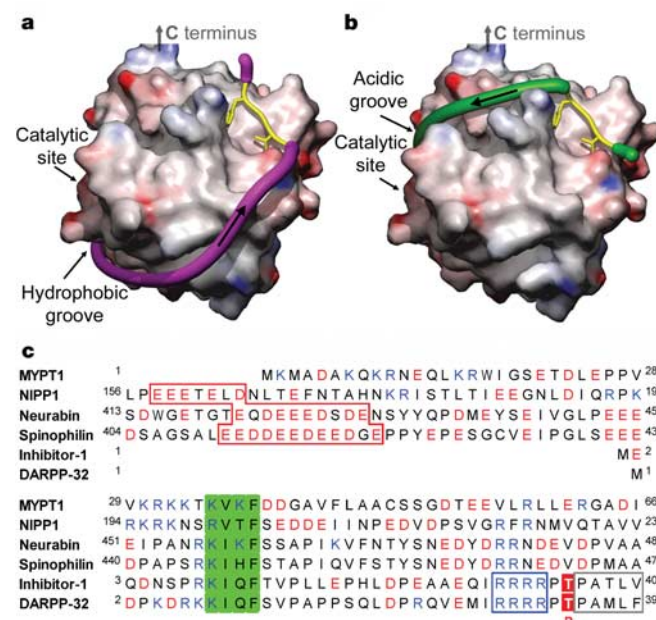


Figure 4 Model illustrating two possible paths between the RVxF motif and the catalytic site of PP1. **a**, Path N-terminal to the RVxF motif (magenta) suggested by the structure of PP1-MYPT1. NIPP1, neurabin and spinophilin may follow this path. **b**, Path C-terminal to the RVxF motif (green) leading to the cleft by way of the acidic groove. Inhibitor-1 and DARPP-32 may follow this path in order to position the inhibitory P-Thr 35 at the catalytic site, allowing for favourable electrostatic interactions between the basic sequence N-terminal to P-Thr 35 and the acidic groove of PP1. **c**, Sequences of some PP1 regulatory subunits around the RVxF motif (shaded green). Acidic and basic sequences that may potentially interact with PP1 are outlined in red and blue, respectively. Inhibitory phosphorylation sites are shaded red.

Thr 161 and Ser 204, and the consensus RVxF motif, which is preceded by a basic sequence as in MYPT1. A highly acidic sequence (EEETELD) 35 amino acids N-terminal to the RVxF motif coincides with one of the phosphorylation sites. This sequence could possibly target the highly basic metal cluster in the catalytic site of PP1. Also consistent with the current structure is an alternative path that leads to the catalytic cleft C-terminal to the RVxF motif (Fig. 4). As previously proposed²⁵, inhibitor 1 and the dopamine and cAMP-regulated phosphoprotein DARPP-32 may follow this second path.

The structure of PP1-MYPT1 presented here reveals how the binding of a regulatory subunit adapts the catalytic cleft of PP1 to perform a specific function. In this case, that function is the dephosphorylation of myosin that leads to smooth muscle relaxation. The interactions involving the RVxF motif play an important part in the formation of the complex. However, other weaker interactions modulate PP1 activity and isoform specificity. We predict that features revealed by this structure will apply to other PP1 regulatory subunits. □

Methods

Protein preparation

The gene encoding chicken gizzard PP1δ (UniProt P37140) was inserted into the pTYB12 IMPACT expression vector from New England BioLabs that leads to expression of the fusion protein chitin-binding domain-intein-PP1δ. The gene encoding MYPT1₁₋₂₉₉ (UniProt Q90624) was inserted into the expression plasmid pACYC184, which is compatible with pTYB12, and placed under the control of a T7 promoter. The two vectors were introduced into *Escherichia coli* strain BL21 (DE3) for co-expression. The complex PP1δ-MYPT1₁₋₂₉₉ was purified using a chitin affinity column. Because expression levels were higher for MYPT1₁₋₂₉₉ than for PP1δ, the affinity tag was added to the latter, to ensure that a 1:1 complex was obtained during purification. After elution from the chitin column, following dithiothreitol-induced self-cleavage of the intein, the complex was loaded into a MonoQ column (Pharmacia) and eluted using a 150 to 400 mM gradient of NaCl in 20 mM Tris-HCl buffer pH 7.5, 1 mM MnCl₂, 2 mM dithiothreitol.

Crystallization

Before crystallization the PP1δ-MYPT1₁₋₂₉₉ complex was dialysed overnight against 10 mM Tris buffer pH 7.5, 20 mM NaCl, 4 mM dithiothreitol and 1 mM MnCl₂ and concentrated to ~8 mg ml⁻¹. The complex was crystallized at 20 °C using the hanging-drop vapour diffusion method by mixing 2 µl of protein solution and 2 µl of a precipitant solution, containing 9.2% (w/v) poly(ethylene glycol) (PEG) 5,000 monomethylether, 35% (v/v) glycerol and 200 mM NH₄Cl.

Data collection, structure determination and refinement

A data set (to 2.7 Å resolution) was collected at 100 K at BioCARS beamline 14-BM-C from a crystal of PP1δ-MYPT1₁₋₂₉₉. The diffraction data were indexed and scaled using the programs DENZO and SCALEPACK²⁶ (Supplementary Table 1). A molecular replacement solution for the PP1δ part of the structure was obtained using the program AMoRe²⁷ and the structure of PP1α (Protein Data Bank code 1FJM) as a search model. After screening a number of ankyrin-repeat-containing structures, a molecular replacement solution was found for a group of three repeats derived from the structure of P18-Ink4C (Protein Data Bank code 1IHB). An electron-density map was calculated with phases derived from the partial model resulting from the molecular replacement solutions of PP1 and the three ankyrin repeats. The map revealed the location of another group of three ankyrin repeats. Additional refinement led to the identification of two additional, non-canonical repeats (repeats 4 and 8). Refinement and model-building were done using the programs CNS²⁸ and O²⁹. The quality of the refined structure was validated with the programs PROCHECK³⁰ and CNS. The final structure includes PP1δ amino acids A1-A309 (chain A), MYPT1 amino acids B1-B291 (chain B), two Mn²⁺ ions, a molecule of PEG 5,000 monomethylether and 228 water molecules. Notice that PP1 residue His A1 is the last of three amino acids left over by the intein construct (the other two being disordered in the structure). The PP1δ chain was numbered starting from this residue to preserve the same notation as for PP1α and PP1γ, which have one additional amino acid at their N termini.

Received 30 January; accepted 16 April 2004; doi:10.1038/nature02582.

Published online 26 May 2004.

1. Bollen, M. Combinatorial control of protein phosphatase-1. *Trends Biochem. Sci.* **26**, 426-431 (2001).

- Cohen, P. T. Protein phosphatase-1—targeted in many directions. *J. Cell Sci.* **115**, 241-256 (2002).
- Hartshorne, D. J., Ito, M. & Erdodi, F. Myosin light chain phosphatase: subunit composition, interactions and regulation. *J. Muscle Res. Cell Motil.* **19**, 325-341 (1998).
- Johnson, D. E. et al. Identification of protein-phosphatase-1-binding domains on the glycogen and myofibrillar targeting subunits. *Eur. J. Biochem.* **239**, 317-325 (1996).
- Hirano, K., Phan, B. C. & Hartshorne, D. J. Interactions of the subunits of smooth muscle myosin phosphatase. *J. Biol. Chem.* **272**, 3683-3688 (1997).
- Tanaka, J. et al. Interaction of myosin phosphatase target subunit 1 with the catalytic subunit of type 1 protein phosphatase. *Biochemistry* **37**, 16697-16703 (1998).
- Toth, A. et al. Study of the subunit interactions in myosin phosphatase by surface plasmon resonance. *Eur. J. Biochem.* **267**, 1687-1697 (2000).
- Gailly, P. et al. Regions of the 110-kDa regulatory subunit M110 required for regulation of myosin-light-chain-phosphatase activity in smooth muscle. *Eur. J. Biochem.* **239**, 326-332 (1996).
- Velasco, G., Armstrong, C., Morrice, N., Frame, S. & Cohen, P. Phosphorylation of the regulatory subunit of smooth muscle protein phosphatase 1 M at Thr 850 induces its dissociation from myosin. *FEBS Lett.* **527**, 101-104 (2002).
- Ichikawa, K. et al. Interactions and properties of smooth muscle myosin phosphatase. *Biochemistry* **35**, 6313-6320 (1996).
- Goldberg, J. et al. Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* **376**, 745-753 (1995).
- Maynes, J. T. et al. Crystal structure of the tumor-promoter okadaic acid bound to protein phosphatase-1. *J. Biol. Chem.* **276**, 44078-44082 (2001).
- Kita, A. et al. Crystal structure of the complex between calyculin A and the catalytic subunit of protein phosphatase 1. *Structure* **10**, 715-724 (2002).
- Egloff, M. P., Cohen, P. T., Reinemer, P. & Barford, D. Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate. *J. Mol. Biol.* **254**, 942-959 (1995).
- Egloff, M. P. et al. Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phosphatase 1. *EMBO J.* **16**, 1876-1887 (1997).
- Connor, J. H., Kleeman, T., Barik, S., Honkanen, R. E. & Shenolikar, S. Importance of the beta12-beta13 loop in protein phosphatase-1 catalytic subunit for inhibition by toxins and mammalian protein inhibitors. *J. Biol. Chem.* **274**, 22366-22372 (1999).
- Sedgwick, S. G. & Smerdon, S. J. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311-316 (1999).
- Kissinger, C. R. et al. Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature* **378**, 641-644 (1995).
- Feng, J. et al. Inhibitory phosphorylation site for Rho-associated kinase on smooth muscle myosin phosphatase. *J. Biol. Chem.* **274**, 37385-37390 (1999).
- Helps, N. R., Barker, H. M., Elledge, S. J. & Cohen, P. T. Protein phosphatase 1 interacts with p53BP2, a protein which binds to the tumour suppressor p53. *FEBS Lett.* **377**, 295-300 (1995).
- Long, X., Wu, G., Gaa, S. T. & Rogers, T. B. Inhibition of protein phosphatase-1 is linked to phosphorylation of p53 and apoptosis. *Apoptosis* **7**, 31-39 (2002).
- Gorina, S. & Pavletich, N. P. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science* **274**, 1001-1005 (1996).
- Beullens, M. et al. Molecular determinants of nuclear protein phosphatase-1 regulation by NIPP-1. *J. Biol. Chem.* **274**, 14053-14061 (1999).
- Terry-Lorenzo, R. T. et al. The neuronal actin-binding proteins, neurabin I and neurabin II, recruit specific isoforms of protein phosphatase-1 catalytic subunits. *J. Biol. Chem.* **277**, 27716-27724 (2002).
- Barford, D., Das, A. K. & Egloff, M. P. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 133-164 (1998).
- Otwiński, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326 (1997).
- Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157-163 (1994).
- Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921 (1998).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallogr. A* **47**, 110-119 (1991).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements R.D. and T.T. were supported by NIH grants. Use of the Advanced Photon Source was supported by the US Department of Energy. The NIH supported use of the BioCARS facilities. We thank B.-J. Gong for assistance with protein expression during the early stages of this work, T. Kitazawa for discussions and A. Borus for reading the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.D. (dominguez@bbri.org). Coordinates have been deposited in the Protein Data Bank under accession number 1S70.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejp.com

naturejobs

Mental barriers

At Stanford University in Palo Alto, California, the Clark Center, the hub of the university's Bio-X multidisciplinary programme, sits just a stone's throw from the medical school, the computer science building, the chemistry department and the engineering department. Does that naturally encourage interaction? Not necessarily, said Charles Kruger, emeritus dean of research and graduate policy, speaking to a delegation from the Biotechnology Industry Organization's annual meeting in San Francisco last week.

"Walking back and forth is no problem," Kruger told the visitors. "The problem is thinking you can do that." Kruger's words resonate at a time when 'interdisciplinary' is a buzzword, but making it really happen is more difficult. The Clark Center, with its central circular courtyard, corridors on the outside, glass-walled labs, and a coffee shop on the top floor, is engineered to create interactions.

Bricks, mortar and style are being used as a tool to recruit people (*Nature* **424**, 858–859; 2003) and facilitate interaction and inspiration (*Nature* **424**, 718–720; 2003). But mindset may be more important than architecture. "How do you get promoted if you are one of many authors on a paper?" asked one audience member. "Who gets credit for a patent when a large team is involved in the discovery" asked another.

Kruger's reply was that people who opt to work in an interdisciplinary environment self-select for cooperation and tend to sort out such issues with a little thought. "The physical barriers here are not as great as the mental barriers," he said. The real issue may be finding such common ground — and career advancement rewards — for scientists who don't work in such environments.

Paul Smaglik
Naturejobs editor



Contents

POSTDOCS & STUDENTS

High-flying postdocs p786

CAREER VIEW

Scientists & Societies

When your spouse gets a job

Graduate Journal

Thrifty solutions

Movers

Ruedi Aebersold

p788

WWW.NATUREJOBS.COM

Career centre

Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Stars in the making

Faster than a speeding centrifuge... able to write grant proposals in a single bound... it's Superpostdoc! Kendall Powell tracks down these all-action figures.

DOM HAMERMAN

Twenty years ago, David Page found a unique DNA probe for the human X and Y chromosomes. He was working on an early version of the human genome map in David Botstein's lab at the Massachusetts Institute of Technology (MIT) and was about to graduate from medical school. Beckoned by the discovery's research potential, Page became one of the first fellows of the Whitehead Institute for Biomedical Research, across the street in Cambridge, Massachusetts. These new positions — or 'superpostdocs' — offered more independence and salary than a traditional US postdoctoral fellowship in the biomedical sciences, but less responsibility and pressure than a junior faculty spot, and without the long-term commitment of a tenure-track position.

Last year, Page, now associate director of science at the Whitehead, grabbed headlines with the genetic map of the human Y chromosome (*Nature* **423**, 825–837; 2003). That work can be traced back to the DNA probe that Page took with him to his fellowship and used to launch his career as a reproductive geneticist. But he admits that doing a superpostdoc was not without its challenges.

As he was setting up his solo laboratory at the age of 28, he watched senior scientists he ostensibly had to compete with move into a new building complete with a "bevy of postdocs" and felt distinctly underqualified for the task. But his tenacity in setting up independent shop proved a "whopping advantage" when he later became an assistant professor.

The Whitehead programme has since been duplicated at universities and institutes around the world, giving promising researchers a chance to bypass the traditional postdoctoral fellowship carried out under an adviser's tutelage. Superpostdocs encourage young scientists to pursue risky, innovative and interdisciplinary ideas. The programmes offer different levels of independence, support and perks, but most give postdocs the chance to set up their own 'mini lab' to pursue an independent line of research.

KRYPTONITE PRESSURE

But along with the autonomy and higher funding come great expectations. "The hurdle is higher indeed" for superpostdocs, says Pierre Wiltzius, director of the Beckman Institute at the University of Illinois, Urbana-Champaign. He says these researchers are expected to be go-getters in return for the hefty investment.

The institute — which fosters multidisciplinary research in biological intelligence, human-computer interaction, and nanostructures — offers three-year fellowships to new PhDs. Applicants propose research plans bringing together teams across the diverse faculty of cognitive scientists, chemists, computer scientists and electrical engineers. Fellows receive an annual salary of \$50,000 and modest start-up funding

to run their research programme either in their own space or within a faculty member's lab.

Beckman fellow Jesse Spencer-Smith, a mathematical psychologist, says that his work on the role of dynamics in human facial expressions requires collaboration with the different traditions represented on campus. Being the first to pursue a new line of research would have been extremely difficult in a conventional postdoc.

The fellowships help some researchers make a transition to new fields. After completing his PhD on the biophysics of bacterial flagella, William Ryu did a one-year postdoc in his graduate adviser's lab at Harvard University in Cambridge, Massachusetts. But when he decided to investigate the behaviour of the



Jesse Spencer-Smith (left) studies the dynamics of facial expressions.

roundworm *Caenorhabditis elegans*, he took a five-year stint at the Lewis-Sigler Institute for Integrative Genomics at Princeton University in New Jersey, where his group strives to quantify the worm's response to temperature. There he has the funding and freedom from grant-writing to pursue high-risk work without having to worry about funding or tenure.

Alan Jasanoff, a current Whitehead Fellow, says he gets to see a side of academia and to establish collaborations that are not open to most postdocs. This has helped to advance both his neural-imaging research and his career. Later this year, when his postdoc ends, he will start as an assistant professor at MIT with a dual appointment in nuclear engineering and brain and cognitive sciences.

FULL SPEED AHEAD

Although most programmes are non-renewable, it is fairly common for superpostdocs to be offered jobs at their fellowship institution. Spencer-Smith will start as a faculty member at the Beckman Institute in August, and Anita Sil, a microbial geneticist, only had to move her lab across the hall from her power postdoc when



SUPERMAN: CORBIS SYGMA

High-flying superpostdocs punch through the barriers to gain their independence.

she joined the faculty at the University of California, San Francisco, in 2003. Inevitably, if positions are available, institutes end up holding on to promising researchers they have fostered. Other superpostdocs have found good research jobs in industry.

Setting up a lab, handling administration and managing people can be daunting for a young scientist fresh out of postgraduate studies. When Sil started her fungal-pathogenesis project, she chose to set up her own lab, even with the hassles of buying glassware, because the research would “clearly be mine and mine alone”, she says. But she admits that for other postdocs, working in an already established lab may be the best path for doing creative experiments quickly.

Two of the biggest challenges, say the research fellows, are competing with senior labs in their field and feeling as if they are in limbo — half postdoc, half junior faculty — without an extensive support network. But during her subsequent job search, Sil says, it was obvious to prospective employers that she was already an independent investigator.

ACCELERATED TRACKS

In mathematics and theoretical physics, one of the most coveted postdoctoral spots is the Institute for Advanced Study in Princeton. Members, or postdocs, work at essentially the same level as assistant professors, explains theoretical physicist Harlan Robins. They start on a salary close or equal to that of an assistant professor, enjoy subsidized housing and a private office (there are no labs), and have free rein for their research.

It is almost impossible for experimental physicists to set up their own ‘mini-labs’, given the extent of collaboration and the equipment needed to make a particle-physics experiment fly. Promising graduate students often apply for postdoctoral fellowships at US national or international laboratories. Two years ago, Argonne National Laboratory began an international competition for outstanding postdocs that carries an

annual \$70,000 stipend and \$20,000 expense account for posts lasting two or three years. Last year, Sandia National Laboratories launched the President Harry S. Truman Fellowship, which offers similar autonomy for postdocs in physics and engineering.

EUROPEAN EXPERIENCE

Recognizing the need to attract talented young scientists to establish their careers in Europe, the European Science Foundation in Strasbourg started the European Young Investigators Awards programme last year. Researchers of any nationality with 2 to 10 years’ postdoctoral experience can apply for up to €250,000 (US\$306,000) a year for five years. Applicants must first find a host institution in one of 15 participating European countries to back their proposal.

Dominique Martin-Rovet, ESF science policy adviser, envisions a “Nobel prize for the younger scientist”. The programme creates positions in which young European scientists can be autonomous, and seems to be a hit with both local and expat European researchers, bringing in 770 applications for the 25 awards to be given out this autumn.

Postdocs hoping to step up to superstar status can also apply for a Career Development Award from the UK Medical Research Council (MRC). This gives experienced postdocs resources to establish themselves as independent researchers with a total budget of about £500,000 (US\$917,000) over four years. It also allows for up to two years of overseas training and career development.

“What we are looking for is the potential for future research leadership,” says Shabih Syed, manager of the research career awards group at the MRC. “We can only look for signs of potential, but we generally get it right.” Judging by the calibre and career paths of superpostdocs, most other programmes do, too.

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

Web links

Whitehead Institute
www.wi.mit.edu/who/whofellows.html
 Beckman Institute
www.beckman.uiuc.edu/fellows/index.html
 Lewis-Sigler Institute
www.genomics.princeton.edu/topics/institute2.html
 Institute for Advanced Study
www.ias.edu/
 ESF
www.esf.org
 MRC
www.mrc.ac.uk
 Argonne National Laboratory
www.anl.gov
 Sandia National Laboratory
www.sandia.gov/employment/employment/special/truman

GRADUATE JOURNAL

A little Canadian goes a long way

Apart from the obvious win in landmass size, Canada has a smaller population, smaller economy, smaller-sized buffets and much smaller military and research budgets than our big brother to the south. However, Canadian pilots pull off Top Gun wins, and Canadians continue to forge ahead in their labs with MacGyver-like ingenuity.

Lord Rutherford, who graced the physics labs of McGill University at the turn of the century, is credited with the battle cry: "We haven't the money, so we've got to think." I would add, "and get out the duct tape". My master's thesis setup used duct tape to retrofit an old pump to avoid buying a new one. Fellow Canadian researcher Paul Campbell cooked up a unique mouse stereotaxic-mount-cum anaesthesia box that incorporated a latex glove with a hole cut out. He admits that a bit of gas leaked into the room, but argues "that's why you start with extra brain cells, isn't it?"

Walking to the local hardware store where they retooled the dusty pump I found at the back of the lab, dreaming up high-tech/low-cost solutions, staying with local friends at foreign conferences — the eternal Canadian search for thrift in research adds yet another educational angle to my graduate student experience.

Sidney Omelon is a PhD student in bone biomaterials at the Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.

SCIENTISTS & SOCIETIES

Two-body problem

When my spouse got an offer in January to teach at a small college, the biggest advantage had to be the location. Ninety minutes from Princeton, Philadelphia and New York, the number of biotech companies in the area was staggering — key for me because I am opting for industry. So, after she accepted the position in January, all we had to worry about was finding me a job.

Several months, one phone interview and one pseudo-offer later, I am still networking away. Making phone calls, sending e-mails, calling in favours and blanketing the area with my resumé has so far landed me little more than compliments on my CV. And puzzlement from academic colleagues. "You would be a great asset to any biotech company," they say. "Why don't you have a job yet?"

This reaction shows how little academics know about how to get into a career in biotech. Luckily, my institution has a great careers centre. And my experience with the Washington University Postdoc Association has exposed me to industry contacts and career advice. So over the past few months, I've been educating myself.

I soon realized that a biotech company is first and foremost a business. My contacts at the bigger companies tell me that interviewing and hiring moves slowly. And it's influenced by many factors — often nothing to do with science. For all I know, I may be in consideration for a job I applied for three months ago. Because of the number of applicants, you rarely hear back from a company unless they are interested in you.

Also, companies hire to fill very specific needs. If you are working in

an area that pharma is interested in at the moment, you are more likely to get noticed. Throughout my career I have made an effort to diversify my training. So I learned X-ray crystallography in graduate school, NMR spectroscopy at my first postdoc, and am currently working in a microbiology department. But most industry PhD positions want individuals with deep training in one area. For those jobs, my broad range of expertise could be seen as unfocused.

So the move date is approaching, and my spouse and I are now considering living in different cities until I find something. My advice to postdocs looking for a real job? Start looking at least a year before you have to leave.

Jeramia Ory is acting president of the Washington University Postdocs Association and a fellow in the Department of Molecular Microbiology, Washington University School of Medicine.

MOVERS Ruedi Aebersold, Professor of Systems Biology, University of Zurich and the ETH



When proteomics pioneer Ruedi Aebersold moves to Zurich this November, it will be both a homecoming to his native Switzerland and a departure — as he leaves the Institute for Systems Biology in Seattle, Washington, although he will retain a research group there.

But the new position will allow Aebersold to stick to his multidisciplinary roots. As chair of systems biology in a new joint program between the Swiss

Federal Institute of Technology (ETH) in Zurich and the University of Zurich, he will contribute to a multidisciplinary environment much like the one at the Institute for Systems Biology, where a life scientist can regularly work with an astrophysicist or an information technology specialist. Aebersold wants to nurture an environment that will drive both technology and science forward.

The two have gone hand-in-hand for Aebersold ever since he tried to sequence proteins more or less manually for his thesis project: "a painful, tedious and slow process," he says. So after his PhD he jumped at the chance to work at Caltech, where Leroy Hood was developing faster and more sensitive ways to address the same problem.

After four years, Aebersold left for a tenure-track post at the University of British Columbia, in Vancouver, Canada, where he continued to improve protein purification while mass spectrometry for

protein analysis — which requires very pure proteins — took off. But when Hood moved to the University of Washington, Seattle, Aebersold jumped at the opportunity to be reunited with his mentor. Years later Aebersold had another tough decision — leave the university with Hood to co-found the Institute of Systems Biology, or stay put. "Once you're in a tenured position at a great university it's not so easy to walk away," Aebersold says.

But the leap proved worthwhile, as it enabled him to work with computational specialists who are fully integrated into the research plans — not a separate support arm. "It's the way many research institutions will have to go," Aebersold predicts. Working with different groups "shakes up" your thinking and helps create new connections, he says. It also prevents you from getting too enamoured of a particular technique or problem.

CV **2000:** Co-founder, Institute for Systems Biology, Seattle, Washington
1993–2000: University of Washington, Seattle: professor of molecular biotechnology (1998–2000) associate professor (1993–98);
1988–93: Assistant professor of biochemistry, University of British Columbia, Vancouver, Canada
1984–88: California Institute of Technology (1984–86 postdoc, 1986–88, senior research fellow)
1984: PhD in cell biology, University of Basel